Modified gravity and cosmology on the largest scales

Alexandra Terrana

A Dissertation submitted to the Faculty of Graduate
Studies in partial fulfillment of the requirements for the
degree of Doctor of Philosophy

Graduate program in Physics and Astronomy

York University

Toronto, Ontario

May 2017

Abstract

This dissertation presents a series of my contributions to research in theoretical cosmology, focusing on aspects of the very large scale universe, particularly dark energy, cosmic acceleration, modified gravity, and cosmic variance. Following an overview of the current understanding of the standard cosmological model in chapter 1, three pertinent topics are discussed in detail. A common theme among all chapters is the desire to explain the properties of the universe on the largest scales.

One of the biggest mysteries on large scales is the need for dark energy to explain the observed accelerated expansion of the late universe. The unsatisfying explanation offered by the standard cosmological model and the associated enormous fine tuning problem have driven considerable interest in infrared (long-distance) modifications of general relativity. In this work, we consider a particularly well motivated modified theory, *massive gravity*, in which the modification is to simply assume that the particle mediating the gravitational force has a non-zero mass. For a mass on the order of the Hubble constant, this theory offers an alternative explanation of the accelerated cosmic expansion. Chapter 2 lays the theoretical groundwork for massive gravity, summarizing its history and formalism.

A fundamental challenge for any modified gravity theory is sequestering the modification to large enough distance scales, so that the predictions match general relativity on solar system scales where it has been tested to high precision. Chapter 3 provides a detailed analysis of massive gravity's ability to screen its extra degrees of freedom, allowing for continuity with general relativity on short distance scales. Further, in chapter 4, we explore the cosmological production and propagation of gravitational waves in an extension of massive gravity, *bigravity*, determining whether there may be any testable deviations from general relativity. Understanding these predictions is crucial, as there is now a vigorous observational program to probe possible deviations from our standard model.

As rapid progress in observational cosmology unfolds, not only is it paramount to construct viable modified gravity theories to test against general relativity, it is necessary to explore which observational methods will be the most powerful for constraining them. This dissertation contains progress on both of these fronts: analyzing potential modified gravity theories, and analyzing potential novel observational probes of the large scale universe. Chapter 5 provides the theoretical framework for one such novel probe, the *large scale kinetic Sunyaev-Zeldovich effect*. This effect is particularly intriguing because of its ability to overcome cosmic variance, and thus help us unlock the secrets of the universe on the largest scales.

Acknowledgments

I am most thankful for the guidance and support provided by my supervisor, Matthew Johnson. I have learned so much from him. He has been a mentor and role model with respect to how to be a good researcher, but even more so, how to be a good person. I sincerely appreciate all of his advice concerning research and careers, and all he has done to make my experience here stimulating and enjoyable. It was an honor to be his first Ph.D. student!

It has been a privilege to discuss and work with many outstanding physics faculty and students at York University and Perimeter Institute. I am truly grateful for productive collaborations with Richard Brito, Vitor Cardoso , and most recently, Anne-Sylvie Deutsch and Moritz Munchmeyer.

I must also thank my wonderful family and friends for continual encouragement throughout this pursuit. I am especially thankful to Fil for emotional support, help with all computer related issues, and inspirational work ethic.

I gratefully acknowledge the funding sources that made my doctoral work possible. I was supported by the Ontario Graduate Scholarship for the first year and the Vanier Canada Graduate Scholarship for the last three years. This research was supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science. Some results presented here were obtained using the Healpix package (`https://sourceforge.net/projects/healpix/`), the Cosmicpy package (`http://cosmicpy.github.io/`), and the LAMBDA web-based CAMB interface (`https://lambda.gsfc.nasa.gov/toolbox/tb_camb_form.cfm`).

Table of Contents

List of Tables

List of Figures

Citations to Published Work

This dissertation includes material from work that has been previously published as follows.

- Chapter 3 has been published in Physical Review D [1].
- Chapter 4 has been published in Physical Review D [2].
- Chapter 5 has been published in the Journal of Cosmology and Astroparticle Physics [3].

Permission from coauthors has been granted for these works to be included in this dissertation. Minor changes to the publications have been implemented here for the sake of consistent notation throughout the dissertation.

Chapter 1

Our universe

Describing the composition and evolution of our universe is the goal of modern cosmology, in which the laws of physics are applied to the largest scales. In the last few decades, cosmologists have made dramatic progress unveiling the mysteries of the universe, transforming the field from a set of speculative theories to a testable precision science.

Continual improvements in technology have enabled us to collect cosmological measurements extending all the way to our observable horizon. Through observations of the cosmic microwave background (CMB), most notably with the Planck satellite [4], we can probe the oldest light in the universe with remarkable precision. This radiation contains a wealth of information about the early universe, its dynamics, and its composition. Additionally, deep galaxy redshift surveys such as the Sloan Digital Sky Survey [5] have mapped millions of galaxies, providing insight towards the distribution and constituents of large scale structure.

In addition to the many observational efforts, there have also been breakthroughs in the theoretical understanding of physical processes that have culminated in a consistent theoretical framework to understand these observations. All measurements to date have led cosmologists to converge on a standard cosmological model, ΛCDM, which appears to be an excellent fit to nearly all the data with just six parameters. We have a model through which we can make predictions, and the technology to conduct experiments through which these predictions can be tested.

Alongside its successes, ΛCDM contains some parameters that are intrinsically dissatisfying. The recent decade of experiments in cosmology has tended to confirm the model, including its dissatisfying aspects, namely the apparent dominance of two unknown quantities: dark matter and dark energy. This introductory chapter highlights the fundamental features of ΛCDM to provide an overview of the current understanding of cosmology.

1.1 Previewing the Cosmological Model

Let us begin by outlining the main discussion points of this chapter and how they are connected to subsequent chapters. The cosmological model can be broken down into two main ingredients:

- The first ingredient pertains to the physical forces describing the interactions and evolution of matter and energy. On such large scales, gravity is the only fundamental force that plays a role. In the standard cosmological model, the force of gravity is described by Einstein's theory of general relativity (GR), in which gravity is essentially the geometry of spacetime. Einstein's revolutionary discovery of GR has been instrumental throughout the last century, forming the foundation for a testable model of the universe. The next section 1.2 provides an introduction and overview of GR and its applications to cosmology.
- The second is the composition of matter and energy in the universe, which are typically categorized as baryons, dark matter, dark energy, and radiation. An important aspect of our theory is identifying the components that dominate the energy budget at various epochs. This is formalized in section 1.3. The standard model, called ΛCDM, is in fact named after the most abundant forms of matter and energy in the universe: Λ represents dark energy described by a cosmological constant, and CDM stands for cold dark matter. These dark constituents cause us to cast doubt on our understanding of the universe under the standard model. They are discussed in section 1.4.

Working together with these ingredients is a key feature in contemporary models: *the cosmological principle*. The principle is based on the idea that the universe is essentially the same everywhere: there is no special location or special direction. On the scale of planets, galaxies, and other dense structure, this simply doesn't hold. It is meant only to apply on the very largest scales. This is one of many cosmological statements that is statistical in nature, as the distribution of over-densities and under-densities, in which local variations are averaged over, is approximately the same in all regions of the universe. Stated in other words, the universe is statistically homogeneous. We'll see how this is built into the theory in section 1.3.1.

To understand the formation of structure in the universe, we have to go beyond the perfectly homogeneous description above. Section 1.3.2 describes how an inflationary phase of accelerated expansion in the very early universe can quantum mechanically generate tiny inhomogeneities on top of a smooth background. Following, in section 1.3.3, we describe cosmological perturbation theory, which is an essential formalism for evolving these initial perturbations into large-scale structure.

The last section of this introduction chapter 1.6, discusses the most important observational probe of large scales in our universe: the cosmic microwave background (CMB). We describe how the CMB provides precise measurements of the composition and geometry of the universe, playing a key role in the development of the cosmological model.

Connection to other chapters

The bulk of this dissertation, appearing in chapters 2-5, relies on the foundation built in this chapter. All chapters can be unified by the goal of understanding the properties of the universe on large scales, which of course heavily depends on understanding the standard cosmological model presented here.

The introduction to GR presented in 1.2 is not only useful as a key component of the cosmological model, but is especially important for the work presented in chapters 2-4, which all focus on modifications to standard GR. These modifications are motivated by the mystery of dark energy and cosmic acceleration as introduced in section 1.4.2, potentially indicating that GR may not be reliable on cosmological scales. This work focuses on a particular category of infrared modified gravity theories called massive gravity in which a non-zero graviton mass offers an alternative explanation for cosmic acceleration. A brief introduction in section 1.5 is supplemented by a thorough review of theoretical aspects of massive gravity in chapter 2, followed by specific applications in chapters 3 and 4.

The work in chapter 4 considers cosmological tests of bigravity, an extension of massive gravity. The aim is to study the effects of the tensor perturbations in bigravity, and derive the prediction for cosmological observables in comparison to the prediction given by GR. This will rely on applying the Friedmann equations introduced in 1.3.1 to find viable background cosmologies. We also analyze the production of tensor perturbations produced during inflation in bigravity, which builds from the discussion of inflation in 1.3.2.

Chapter 5 switches gears slightly to analyze observational probes of large scales. In particular, we focus on the large scale kinetic Sunyaev-Zeldovich effect, which produces secondary anisotropies in the CMB. A basic understanding of the CMB as presented in 1.6 will be essential for this work, as will the fundamentals of cosmological perturbation theory introduced in 1.3.3.

1.2 General Relativity

Einstein's general relativity (GR) describes gravitation through the differential geometry of curved spacetime. GR is an intrinsic part of modern physics and lies at the heart of the standard cosmological model. Here, we outline the most important concepts required to understand cosmology, namely the spacetime metric and the Einstein equations. Applying the Einstein equations to the Friedmann-Robertson-Walker (FRW) metric yields the fundamental relationships between the metric parameters and the density components in the universe.

This brief section on GR simply covers the basics of the theory as needed to understand cosmology in the next section. Both sections 1.2 and 1.3 overview only the fundamentals, following textbooks such as [6, 7]. Natural units $\hbar = c = 1$ are employed throughout, and the planck mass is used interchangeably with Newton's gravitation constant according to $M_P = 1/\sqrt{8\pi G_N}$.

The spacetime metric

In Einstein's theory, what we experience as gravity is due to the intrinsic curvature of spacetime. Mathematically, spacetime is a manifold whose points correspond to physical events which are represented by four coordinates written as a four-vector, $x^\mu = \{x^0, x^1, x^2, x^3\}$, for which one choice is Cartesian coordinates $\{t, x, y, z\}$.

The metric is the central object in GR that describes the connection between spacetime events. Moreover, the metric specifies all of the geometric and causal structure of spacetime, allowing us to measure distances, times, volumes, and angles. It can be written as a 4×4 symmetric matrix with entries labeled $g_{\mu\nu}$, and is commonly expressed as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \,, \tag{1.1}$$

where dx^μ is an infinitesimal coordinate displacement, and ds is the infinitesimal spacetime interval or line element. Greek indices μ and ν range from 0 to 3 and repeated indices are summed over in the Einstein convention. Cosmologists typically employ the mostly positive Lorentzian metric signature $\{-, +, +, +\}$. The inverse metric is denoted with upper indices and has the property that $g^{\mu\alpha} g_{\mu\beta} = \delta^\alpha_\beta$, where δ^α_β is the Kronecker delta. The metric is a tensor, meaning that it has specific transformation properties under a change of coordinates:

$$g_{\mu\nu} = \frac{\partial x^{\mu'}}{\partial x^\mu} \frac{\partial x^{\nu'}}{\partial x^\nu} g_{\mu'\nu'} \tag{1.2}$$

The spacetime interval ds^2 in (1.1) is invariant under changes of coordinates, called diffeomorphism invariance, meaning that observers in different inertial frames will all measure

the same interval ds. The invariant interval ds^2 captures the causal structure of spacetime. Unlike purely spatial intervals, this spacetime interval can be positive, negative, or zero. When $ds^2 < 0$, the interval is *timelike* and $\sqrt{|ds^2|}$ represents proper time. Only timelike intervals can be physically traversed by massive objects. When $ds^2 = 0$, the interval is *lightlike*, or *null*, and can only be traversed by light. When $ds^2 > 0$, the interval is *spacelike* and ds represents incremental proper length. Spacelike intervals connect events that are outside each other's light cones, and therefore cannot be traversed.

Special relativity is described by the simplest metric, a flat spacetime metric, or Minkowski metric: $g_{\mu\nu} = \eta_{\mu\nu}$. In matrix notation, it is given by

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.3}$$

and therefore the invariant flat spacetime interval is

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2. \tag{1.4}$$

When working in situations with spherical symmetry, it is helpful to use polar coordinates. This corresponds to a change of coordinates $\{t, x, y, z\} \to \{t, r, \theta, \phi\}$, resulting in a different foliation of spacetime. The interval in this case is written as,

$$ds^2 = -dt^2 + dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right). \tag{1.5}$$

Geodesics and Derivatives

Two important sets of equations in GR govern the evolution and interaction of matter, energy, and spacetime. The first is the geodesic equation which describes the paths of particles in the absence of any non-gravitational forces. It is the generalization to curved spacetime of Newton's second law with no forces, $d^2\vec{x}/dt^2 = 0$. The full geodesic equation is

$$\frac{d^2x^\mu}{d\lambda^2} = -\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}, \tag{1.6}$$

where λ is any affine parameter that parametrizes the particle's trajectory, such as proper time, and $\Gamma^\mu_{\alpha\beta}$ are the Christoffel symbols, also known as the metric connection, that capture the curvature of spacetime. They are given by specific combinations of derivatives of the metric,

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2}g^{\mu\nu}\left(\partial_\beta g_{\alpha\nu} + \partial_\alpha g_{\beta\nu} - \partial_\nu g_{\alpha\beta}\right), \tag{1.7}$$

where $\partial_\mu = \frac{d}{dx^\mu}$. The Christoffel symbols also determine the covariant derivative ∇_μ, which can be thought of as the extension of the partial derivative ∂_μ to curved spacetime. This extension

is required since the partial derivative does not transform as a tensor under coordinate transformations on an arbitrary manifold. By correcting the partial derivative with the Christoffel symbols, we obtain a derivative operator that properly transforms as a tensor. For example, the covariant derivative acting on a vector V^ν and tensor $T_{\alpha\beta}$ is

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\sigma} V^\sigma, \qquad \nabla_\mu T_{\alpha\beta} = \partial_\mu T_{\alpha\beta} - \Gamma^\sigma_{\mu\alpha} T_{\sigma\beta} - \Gamma^\sigma_{\mu\beta} T_{\alpha\sigma}. \tag{1.8}$$

The Christoffel symbols satisfy the metric compatibility condition: $\nabla_\alpha g_{\mu\nu} = 0$.

Einstein Equations

The second set of fundamental equations in GR are the Einstein equations, which are the covariant generalization of the Poisson equation for the Newtonian gravitational potential. These equations relate the metric that describes gravitation to the matter and energy in the universe. They are compactly written as

$$G_{\mu\nu} = 8\pi G_N T_{\mu\nu}\,, \tag{1.9}$$

where the left hand side describes the curvature of spacetime in the Einstein tensor, $G_{\mu\nu}$, and the right hand side describes the distribution of matter and energy in the energy-momentum tensor $T_{\mu\nu}$. G_N is the Newtonian gravitational constant which, in our choice of units, is related to the Planck mass as $M_P = 1/\sqrt{8\pi G_N}$. The Einstein tensor is typically expressed in terms of the Ricci tensor $R_{\mu\nu}$ and Ricci scalar R,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}. \tag{1.10}$$

The Ricci tensor and scalar are built from the Riemann tensor, $R^\alpha{}_{\mu\beta\nu}$, which defines spacetime curvature. The Riemann tensor is constructed from second derivatives of the metric, but can also be written in terms of the Christoffel symbols:

$$R^\alpha{}_{\mu\beta\nu} = \partial_\beta \Gamma^\alpha_{\mu\nu} - \partial_\nu \Gamma^\alpha_{\mu\beta} + \Gamma^\alpha_{\sigma\beta}\Gamma^\sigma_{\mu\nu} - \Gamma^\alpha_{\sigma\nu}\Gamma^\sigma_{\mu\beta}. \tag{1.11}$$

The Ricci tensor is given by the contraction $R_{\mu\nu} = g_{\alpha\sigma} g^{\sigma\beta} R^\alpha{}_{\mu\beta\nu} = R^\alpha{}_{\mu\alpha\nu}$, and the Ricci scalar by $R = g^{\mu\nu} R_{\mu\nu} = R^\mu_\mu$.

The Riemann tensor has many symmetry properties in permutations of its indices. An important consequence of these symmetries is a differential identity known as the Bianchi identity:

$$\nabla_\lambda R_{\rho\sigma\mu\nu} + \nabla_\rho R_{\sigma\lambda\mu\nu} + \nabla_\sigma R_{\lambda\rho\mu\nu} = 0. \tag{1.12}$$

Some manipulation of the indices reveals that the twice contracted Bianchi identity implies conservation of the Einstein tensor,

$$\nabla^\mu G_{\mu\nu} = 0. \tag{1.13}$$

From the Einstein equations (1.9), this coincides with the generalization of conservation of energy and momentum in curved spacetime,

$$\nabla^\mu T_{\mu\nu} = 0. \tag{1.14}$$

Although written compactly in equation (1.9), the Einstein equations are extremely complicated non-linear partial differential equations for the metric tensor field $g_{\mu\nu}$. It is very difficult to solve these equations in generality, and so it is necessary to assume that the metric has some degree of symmetry. In the next section 1.3, we'll see an example, the FRW metric, with maximally symmetric spatial slices.

Einstein-Hilbert Action

The Einstein equations given above can be derived from an action principle. In this perspective, GR is a classical field theory in which the dynamical field is the metric tensor itself. This approach is beneficial because it provides a parallel between GR and other classical field theories described by an action. GR remains special, however, since most other field theories rely on a pre-existing spacetime geometry, while the geometry in GR is determined by the equations of motion. The Einstein equations are yielded as the equations of motion from the Einstein-Hilbert action, which is the simplest possible action for the metric,

$$S_{\text{EH}} = \frac{M_P^2}{2} \int d^4x \sqrt{-g} R\,, \tag{1.15}$$

where $g = \det(g_{\mu\nu})$ is the determinant of the metric and R is the Ricci scalar. The full action of the theory contains an additional term to describe any matter and energy fields, denoted by a matter lagrangian,

$$S_{\text{GR}} = \int d^4x \sqrt{-g} \left(\frac{M_P^2}{2} R + \mathcal{L}_{\text{matter}} \right). \tag{1.16}$$

The principle of least action implies that the variation of this action with respect to the metric is zero,

$$\frac{\delta S_{\text{GR}}}{\delta g^{\mu\nu}} = 0. \tag{1.17}$$

This variation requires careful variation of the determinant as well as the Ricci scalar, which is built out of the Riemann tensor, which contains the Christoffel symbols. These calculations reveal that the resulting equations of motion are indeed the Einstein equations (1.9) if we define the relationship between the energy-momentum tensor and the matter lagrangian,

$$T_{\mu\nu} = \frac{-2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L}_{\text{matter}})}{\delta g^{\mu\nu}} = -2\frac{\delta \mathcal{L}_{\text{matter}}}{\delta g^{\mu\nu}} + g^{\mu\nu}\mathcal{L}_{\text{matter}}. \tag{1.18}$$

1.3 Standard Cosmology

1.3.1 The expanding universe with FRW

FRW metric

We would like a metric to describe cosmology that reflects the homogeneity implied by the cosmological principle. The Friedmann-Robertson-Walker (FRW) metric accomplishes this task. It is invariant under spatial translations (homogeneous) as well as under spatial rotations (isotropic), yet still evolving in time. Essentially, the FRW metric is the metric of our entire universe on the largest scales. It is written as,

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{1.19}$$

where $a(t)$ is the dimensionless scale factor which characterizes the size of spatial slices, and K controls the curvature of spatial slices. By convention, the scale factor today is equal to one,[1]

$$a_{\text{today}} \equiv a_0 = 1. \tag{1.20}$$

Symmetries in the metric are apparent by the invariance of the interval under the substitutions $r \to \sqrt{|K|}r$, $a \to a/\sqrt{|K|}$, $K \to K/|K|$. One relevant parameter remains, $K/|K|$, for which there are three possibilities, termed flat, closed, and open:

$$\begin{aligned} K = 0, &\quad \text{flat (no spatial curvature)} \\ K > 0, &\quad \text{closed (positive spatial curvature)} \\ K < 0, &\quad \text{open (negative spatial curvature)} \end{aligned} \tag{1.21}$$

The FRW metric above is written in comoving coordinates. This is a natural coordinate choice since comoving observers, those at constant $\{r, \theta, \phi\}$, will observe the universe to be isotropic. Comoving observers are said to be moving along with the "Hubble flow." The expansion of the universe is thus built into the scale factor, implying that distances between objects change according to the scale factor. The FRW metric describes the expanding universe from the big bang singularity at $a = 0$ to today $a = 1$. The comoving time coordinate, t, is the elapsed time since the big bang according to a comoving observer's clock. A different time coordinate τ can be introduced, called conformal time, defined by,

$$\tau = \int_0^t \frac{dt'}{a(t')}. \tag{1.22}$$

Changing coordinates to conformal time causes the scale factor to become a conformal factor, multiplying the entire metric,

$$ds^2 = a^2(\tau)\left[-d\tau^2 + \frac{dr^2}{1-Kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right]. \tag{1.23}$$

[1] Present day values will always be denoted with a subscript 0.

With the choice of conformal time, null or lightlike trajectories can all be described by $r = r_0 \pm \tau$, hence they travel on 45° lines on a spacetime diagram. Let us also mention a commonly used perturbed form of the FRW metric, referred to as conformal Newtonian gauge. Specializing for now to $k = 0$, and assuming no anisotropic stress, the perturbed metric is written in terms of the Newtonian gravitational potential Ψ of classical Newtonian gravity:

$$ds^2 = a^2(\tau)\left[-(1+2\Psi)d\tau^2 + (1-2\Psi)\left(dr^2 + r^2 d\theta^2 + \sin^2\theta d\phi^2\right)\right]. \tag{1.24}$$

Substituting these metric coefficients into Einstein equations shows that the Newtonian gravitational potential satisfies the non-relativistic Poisson equation $\nabla^2\Psi = 4\pi G_N \rho$ on small scales, where $\rho = T_{00}$ is the mass density. This metric is important for analyzing growth of structure in the universe in cosmological perturbation theory which will be discussed in section 1.3.3.

We would like to apply the Einstein equations to the FRW metric to analyze the implications of GR for cosmology. We first need to think about the right hand side of the Einstein equations which contains the information about matter and energy in the universe. It turns out that most of the ingredients in the universe have a decent effective description as a perfect fluid. The energy-momentum tensor for a perfect fluid is given by

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}, \tag{1.25}$$

where $u^\mu = dx^\mu/dt$ is the four-velocity of the fluid elements, ρ is the mass density, and p is the pressure. The perfect fluid is further specified by an equation of state, a relationship between ρ and p: $p = w\rho$. Two of the most common examples of cosmological fluids are dust (or matter) and radiation. Dust is collisionless, pressureless, nonrelativistic matter, and obeys $w_m = 0$. The equation of state parameter for radiation is $w_r = 1/3$, and is used to describe actual electromagnetic radiation as well as relativistic massive particles, moving at velocities so close to the speed of light that they are effectively indistinguishable from photons.

Assuming that the fluid is homogeneous, we can take the fluid to be at rest in comoving coordinates, in which case the four-velocity is $u^\mu = \{1, 0, 0, 0\}$, and the energy momentum tensor is,

$$T^\mu{}_\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{1.26}$$

Gravitational redshift

Let's explore trajectories of particles in an FRW universe. This will provide insight to the concept of gravitational redshift in cosmology. Here, we apply the geodesic equation for the FRW metric to a massless particle. We start by defining a four-vector for the energy-momentum as,

$$p^\mu = \frac{dx^\mu}{d\lambda} = \{E, \vec{p}\}. \tag{1.27}$$

This is in fact an implicit definition of the affine parameter λ in equation (1.6). It's not important to determine this explicitly, as it can be eliminated through $d/d\lambda = (dt/d\lambda)d/dt$. The zeroth component of the geodesic equation (1.6) becomes,

$$\begin{aligned} E\frac{dE}{dt} &= -\Gamma^0_{\mu\nu}p^\mu p^\nu \\ &= -\delta_{ij}aa'p^i p^j, \end{aligned} \tag{1.28}$$

where we subbed in the non-zero Christoffel symbols calculated from the FRW metric. Since a massless particle satisfies $ds^2 = 0$ in equation (1.1), we can also write,

$$\begin{aligned} 0 &= g_{\mu\nu}p^\mu p^\nu \\ &= -E^2 + \delta_{ij}a^2 p^i p^j. \end{aligned} \tag{1.29}$$

Combining equations (1.28),(1.29) yields

$$\frac{dE}{dt} = -\frac{a'}{a}E \quad \rightarrow \quad \frac{dE}{da} = -\frac{E}{a}\,, \tag{1.30}$$

the solution to which is,

$$E \propto \frac{1}{a}. \tag{1.31}$$

This is an important result: the energy of a massless particle decreases as the universe expands. Since energy is also inversely proportional to wavelength, we see that wavelength is stretched along with expansion. A photon emitted with wavelength $\lambda_{\rm em}$ at scale factor $a_{\rm em} = a(t_{\rm em})$, will be observed at $a_{\rm obs}$ with a longer wavelength $\lambda_{\rm obs}$,

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = \frac{a_{\rm em}}{a_{\rm obs}}. \tag{1.32}$$

Cosmologists refer to this phenomenon in terms of the *redshift*, z, between the two events, defined by,

$$z = \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm obs}}. \tag{1.33}$$

Assuming that the observation takes place today ($a_{\rm obs} = a_0 = 1$), the relationship between scale factor when the photon is emitted and redshift is made clear:

$$a = \frac{1}{1+z}. \tag{1.34}$$

This redshift is different than the conventional Doppler effect, as it is not caused by the relative velocities of the observer and emitter, but by the actual expansion of space. However, astronomers do often think of the redshift in terms of a velocity $v = cz$ where c is the speed of light. Even though it doesn't make sense to compare relative velocities at different points of curved spacetime, this rule of thumb works well over sufficiently short distances.

Dynamics of FRW

Before heading to Einstein's equations, let's see what we can learn from the conservation of energy equation (1.14). The zeroth component reads,

$$\begin{aligned} 0 =& \nabla_\mu T^\mu{}_0 \\ 0 =& \partial_\mu T^\mu{}_0 + \Gamma^\mu_{\mu\nu} T^\nu{}_0 - \Gamma^\nu_{\mu 0} T^\mu{}_\nu \\ 0 =& -\rho' - 3\frac{a'}{a}(\rho + p) \\ \frac{\rho'}{\rho} =& -3(1+w)\frac{a'}{a}\,, \end{aligned} \tag{1.35}$$

where the prime denotes differentiation with respect to coordinate time t. We can instead consider the scale factor as the independent variable, in which case the differential equation is easily solved,

$$\frac{d\rho}{da} = -\frac{3(1+w)}{a}\rho \qquad \rightarrow \qquad \rho \propto \exp\left[-3\int_a^1 \frac{da'}{a'}(1 + w(a))\right]. \tag{1.36}$$

For constant equation of state parameter this reduces to,

$$\rho = \rho_0 a^{-3(1+w)}, \tag{1.37}$$

which describes the evolution of the mass density as a power law with respect to a in terms of the density today ρ_0. For matter and radiation, the power law is,

$$\rho_m = \frac{\rho_{m,0}}{a^3} \qquad \rho_r = \frac{\rho_{r,0}}{a^4}. \tag{1.38}$$

The matter density falls off proportionally to the volume, which is simply interpreted as the decrease in the number density of particles as the universe expands. Radiation falls off faster because, although the number density decreases the same way, there is an additional loss of energy as a^{-1} as they redshift (see equation (1.31)). These relationships are consistent with the fact that the early universe, which was smaller and denser, was radiation dominated. With expansion, radiation has diluted faster than matter, so matter now dominates over radiation with $\rho_{m,0}/\rho_{r,0} \sim 3300$.

Now let's see what the Einstein equations imply for the FRW metric. Substituting the metric (1.19) and energy-momentum (1.26) components into equations (1.9), with $\mu\nu = 00$ and $\mu\nu = ij$ results in,

$$H^2 \equiv \left(\frac{a'}{a}\right)^2 = \frac{8\pi G_N}{3}\sum_i \rho_i - \frac{K}{a^2} \tag{1.39}$$

$$\frac{a''}{a} = -\frac{4\pi G_N}{3}\sum_i (\rho_i + 3p_i), \tag{1.40}$$

where the index i represents the different types of fluid in the universe (matter, radiation, etc). The above two equations (1.39)-(1.40) are called the Friedmann equations.

Before studying the solutions to these equations, an important cosmological parameter should be highlighted. The Hubble parameter, H, defined above in (1.39), characterizes the rate of expansion. Cosmologists often refer to the Hubble radius as H^{-1} and the "comoving Hubble radius" as $(aH)^{-1}$, which is the radius of the comoving Hubble sphere. The value of the Hubble parameter today is referred to as the Hubble constant, H_0, and is conventionally given in units of km/s/Mpc. The Hubble constant has units of $[\text{length}]^{-1} = [\text{time}]^{-1} = [\text{energy}]$ and therefore sets the most important cosmological length, time, and energy scales.

There is currently controversy over the actual value of the Hubble constant, with different measurement techniques yielding values in tension at the $\sim$ 2-3 σ level. For instance, H_0 can be measured from the cosmic microwave background (CMB) in the early universe, as well as from Cepheid variables and supernovae in the late universe. The latest CMB results from Planck measure the Hubble constant to be $H_0 = (67.3 \pm 1.0)$ km/s/Mpc [8], while recent precise local measurements yield a significantly higher measurement, $H_0 = 73.24 \pm 1.74$ km/s/Mpc [9]. In a homogeneous and isotropic universe, these two approaches should yield the same expected value. This tension remains to be resolved, and could be a hint that we need to consider extensions beyond the standard model. Putting this issue aside, we will adopt the value from Planck throughout. In terms of length, time, and energy, this translates to

$$H_0^{-1} = 4458 \text{ Mpc} = 1.4 \times 10^{26} \text{ m} \tag{1.41}$$

$$H_0^{-1} = 14.5 \text{ Gyr} = 4.58 \times 10^{17} \text{ s} \tag{1.42}$$

$$H_0 = 1.4 \times 10^{-33} \text{ eV}. \tag{1.43}$$

It will also be convenient (for chapter 4) to introduce the conformal Hubble parameter using the definition in (1.22),

$$\hat{\mathcal{H}} = \frac{\dot{a}}{a} = a', \tag{1.44}$$

where dot represents differentiation with respect to conformal time τ. Since $a_0 = 1$, the present day Hubble constant and conformal Hubble constant are equal $\mathcal{H}_0 = H_0$.

Using the Friedmann equations and equation (1.35), assuming a flat universe with a single perfect fluid, and a time-independent equation of state, we arrive at

$$a' = H_0 a^{1-3(1+w)/2} \quad \rightarrow a \propto \begin{cases} t^{\frac{2}{3(1+w)}} & \text{for} \quad w \neq -1, \\ e^{H_0 t} & \text{for} \quad w = -1, \end{cases} \tag{1.45}$$

implying that in a matter dominated universe, the scale factor grows as $a \propto t^{2/3}$, and in a radiation dominated universe, the scale factor grows as $a \propto t^{1/2}$. In either case, we have as expected, a universe that is growing monotonically. We could also express this in terms of conformal time,

$$\frac{\dot{a}}{a} = H_0 a^{1-3(1+w)/2} \quad \rightarrow \quad a \propto \begin{cases} \tau^{\frac{2}{1+3w}} & \text{for} \quad w \neq -1, \\ \frac{-1}{\tau H_0} & \text{for} \quad w = -1, \end{cases} \tag{1.46}$$

which gives a growth rate of $a \propto \tau^2$ for matter domination, and $a \propto \tau$ for radiation domination. It is useful to recast the Friedmann equations in terms of an energy budget by defining the density parameter Ω as

$$\Omega = \sum_i \Omega_i = \sum_i \frac{\rho_i}{\rho_{\text{crit}}}, \tag{1.47}$$

where the critical density is defined as,

$$\rho_{\text{crit}} = \frac{3H^2}{8\pi G_N}. \tag{1.48}$$

With these definitions in hand, the Friedmann equation (1.39) is simply

$$1 = \Omega + \Omega_K, \quad \text{with} \quad \Omega_K \equiv -\frac{K}{H^2 a^2}. \tag{1.49}$$

Referring back to equation (1.21), it is evident that the universe is flat for $\Omega = 1$ ($\rho = \rho_{\text{crit}}$), closed for $\Omega > 1$ ($\rho > \rho_{\text{crit}}$), and open for $\Omega < 1$ ($\rho < \rho_{\text{crit}}$).

Connecting with Observations

Let's now use this machinery to connect with observations of our universe. We need to observationally determine a number of parameters: what proportions of matter and radiation are in the universe, what role does curvature play, and what does this mean for the evolution of the scale factor with time?

It used to be thought that our universe contained two components only: matter and radiation. By equation (1.40), and the fact that matter and radiation densities and pressures are positive, we see that $a'' < 0$. Since we know from observations of distant galaxies that the universe is expanding, $a' > 0$, this seems to imply that the universe should be decelerating.

This fits perfectly with our intuition about how gravitational attraction works against the expansion. The greater the mass density in the universe, the more the expansion is slowed by gravity. Additionally, the scaling of ρ_m and ρ_r with scale factor in equation (1.38) agree with observations that our universe was initially radiation dominated, followed by matter domination.

So far, so good, but peculiarities now start to emerge. Let us rewrite the Friedmann equation (1.39) or (1.49) once more in terms of the present values of the density parameters,

$$\frac{H^2}{H_0^2} = \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} - \frac{K}{H_0^2 a^2}, \tag{1.50}$$

This model predicts that the universe should be "curvature dominated" at late times since the curvature term will overwhelm the matter and radiation terms on the right side of the equation. However, our observations tell us that, even though our universe is over 10 billion years old, the K/a^2 term still makes an unobservably small contribution to the Friedmann equations. This is called the flatness problem. Combining data from the Planck satellite with observations of the baryon acoustic oscillations provides a strong constraint on spatial curvature: $\Omega_{K,0} = 0.000 \pm 0.005$ [8]. Remarkably, our universe appears to be extremely flat. From now on, we can safely set $K = 0$. From equation (1.49), this also implies that $\Omega = 1$, and so the total energy density of the universe is equal to the critical density.

With $K = 0$, another inconsistency emerges. There is strong evidence that matter and radiation are not enough, and that there must be some other form of energy density in the universe to reconcile this theory with observations. First of all, we can measure the contributions of matter and radiation to the energy density budget. There are a variety of ways to do this, all of which are in good agreement. Recent estimations of the density parameters yield [8],

$$\Omega_{m,0} = 0.3099 \pm 0.0062, \qquad \Omega_{r,0} \sim 9 \times 10^{-5}. \tag{1.51}$$

To balance the Friedmann equation (1.49), there needs to be another contribution to the energy density to make up for the remaining $\sim 70\%$, $\Omega = \Omega_r + \Omega_m + \Omega_?$. Cosmologists call this mysterious contribution *dark energy*, denoting its density parameter as Ω_{DE}.

Another crucial observation motivates the need for some sort of dark energy. In the 1990's astronomers set out to measure the expansion rate of the universe. As discussed above, it was expected that the universe was decelerating, and the question was, by how much? Measuring the expansion rate requires some sort of cosmological distance measure. With precise information about the dependence of a distance measure on redshift, the effects of geometry and matter content can be disentangled. One such measure is the luminosity distance which is defined in terms of the luminosity L and the flux F of a distant source as

$d_L = \sqrt{L/4\pi F}$. In terms of the comoving distance χ, the luminosity distance is related to the redshift as

$$d_L = (1+z)\chi \tag{1.52}$$

In a flat universe, the comoving distance χ is equal to the FRW coordinate distance r. Its formula comes from considering light from a distant source traveling towards us along a radial null geodesic: $ds^2 = 0 = -dt^2 + a^2 d\chi^2$. Rearranging gives the comoving distance to the source,

$$\chi \equiv \int_{t_{\rm em}}^{t_0} dt \frac{1}{a} = \int_{a_{\rm em}}^{1} da \frac{1}{aa'} = \int_{a_{\rm em}}^{1} da \frac{1}{a^2 H}. \tag{1.53}$$

Using the Friedmann equation to express the Hubble parameter as a function of a, and not yet making assumptions about the equation of state for dark energy $w(a)$ (recall equation (1.36)), yields a geometric expression for the luminosity distance,

$$d_L = \frac{1+z}{H_0} \int_{a_{\rm em}}^{1} da \left[\Omega_{r,0} + a\Omega_{m,0} + a^4 \Omega_{\rm DE,0} \exp\left(-3 \int_a^1 da'(1+w(a'))/a' \right) \right]^{-1/2}. \tag{1.54}$$

Since astronomers typically measure magnitudes, it is useful to convert the luminosity distance to the difference in apparent (m) and absolute (M) magnitudes using

$$\mathrm{m} - \mathrm{M} = 5 \log_{10} \frac{d_L}{10\mathrm{pc}} + \kappa \tag{1.55}$$

where κ is a factor to account for the shift in frequency of the source due to the expansion of the universe. Considering equations (1.54) and (1.55), one sees that if we were able to measure the magnitudes m$-$M and redshift z, then we can test our theory by varying the parameters Ω_i in (1.54). This is best achieved with Type Ia supernovae because their absolute magnitude is always approximately the same (with 10% scatter); they explode at the consistent critical mass given by the Chandraesekhar limit. Type Ia supernovae are therefore called standard candles. Any difference between the apparent brightness of two such supernovae must be a result of their different distances. These objects have been instrumental in allowing astronomers to get accurate distance measurements to distant objects, thus inferring the expansion rate of the universe.

In 1998, results were published from two independent research groups, the High-Z Supernova Search Team [10], and the SuperNova Cosmology Project [11]. Distant supernovae appeared fainter than one would expect if the universe was dominated by matter and radiation alone. This surprisingly suggests that the expansion of the universe is in fact accelerating, not decelerating as expected. The universe must have been expanding more slowly early on, so light had more time to travel from distant objects to us. These distant objects would therefore appear fainter to us than if the universe contained only matter and radiation. The data simply cannot be fit with a matter dominated universe: dark energy is necessary.

Since then, these measurements have been confirmed by numerous independent experiments. Recently, the pan-STARSS survey [12] has provided additional convincing support for accelerated cosmic expansion via supernovae measurements. Furthermore, measurements of the cosmic microwave background, gravitational lensing, and large-scale structure are all consistent. The researchers who led the original discovery were awarded the Nobel Prize in Physics in 2011.

Since we know that gravitation acts to pull matter together, the detection of cosmic acceleration necessitates some form of dark energy that repels gravity and drives the expansion. Referring back to equation (1.40), it is evident that acceleration, $a'' > 0$, requires an energy component with $\rho + 3p < 0$, or $w < -1/3$. This mysterious fluid must have negative pressure. It must also comprise about 70% of the energy density budget.

GR has served us quite well in constructing a framework in which we can model our universe and test the predictions. It has allowed us to link the geometry of spacetime to the mass density of the universe. However, the observational data indicate a need to go beyond standard matter and energy. Section 1.4 elaborates on these mysterious dark constituents, and how we understand them.

1.3.2 Inflation

A discussion of the standard cosmological model would not be complete without *inflation*. This brief section motivates the need for inflation in the very early stages of the universe, and describe its qualitative features.

Based on the crucial observations that the universe is homogeneous, isotropic, and expanding, we can deduce that the universe expanded from a hot and dense early phase where radiation was the dominant contribution to the energy density. This hot big bang model has been rather successful in explaining the results from different observations, but some "puzzles" remain concerning how the radiation era began. The canonical puzzles that motivate the need for inflation are the *flatness problem* and the *horizon problem*. Following a description of these problems, we present the solution of inflation and its standard implementation with a scalar field.

Flatness problem

We've already been introduced to the flatness problem in the discussion following eq. (1.50). Our observations tell us that the present day energy density in curvature is very close to

zero, implying that our universe is quite flat. This also implies that the universe must have been even more flat in its very early stages. To see this, consider equations (1.49) and (1.45). Firstly, eq. (1.45) tells us that for any fluid with $w > -1/3$, a' will decrease as a increases. Therefore, the comoving Hubble radius $(aH)^{-1} = 1/a'$ has been increasing throughout the matter ($w_m = 0$) and radiation ($w_r = 1/3$) dominated eras. Now considering eq. (1.49) and the fact that we measure $\rho \simeq \rho_{\text{crit}}$ today, the density must have been even closer to the critical density in the past ($(aH)^{-1}$ will decrease going back in time, making Ω_K decrease, and Ω very close to 1). The flatness problem is: why was the density so extremely close to the critical density at the start of the radiation era? In other words, how did our universe begin so flat?

Horizon problem

The horizon problem concerns the statistical homogeneity and isotropy of the universe, observed for instance via the CMB (see section 1.6): the CMB temperature is approximately the same in all directions. This is a problem because, given the finite speed of light, it is impossible for the different regions of the universe to have been in contact between the big bang at $t_i = 0$ and decoupling at t_{dec} when the CMB was formed.

This can be formalized by examining the maximum comoving distance $\chi_{\max}$ from which we (today $a_0 = 1$) can receive a photon emitted since the big bang ($a_i = 0$). This measure is helpful since particles that are separated by distances greater than χ could have never communicated with each other. Recalling the comoving distance from equation (1.53) and substituting using eq. (1.45) yields,

$$\chi_{\max} = \int_{a_i}^{a_0} \frac{da}{aa'} = \int_0^1 \frac{da}{H_0 a^{1-(1+3w)/2}}. \tag{1.56}$$

From this expression, we see that $\chi_{\max}$ is finite for $w > -1/3$ near $a = 0$. The finiteness of the maximum comoving distance implies a serious problem: most spots on the CMB sky have non-overlapping past light cones and thus were never in causal contact. We can see this more clearly by relating the size of the horizon at decoupling to an angular scale. Using equations (1.53) and (1.50), and assuming that the universe was radiation dominated ($w_r = 1/3$) before decoupling, it follows that the horizon size at decoupling is

$$\chi_h = \int_{a_i}^{a_{\text{dec}}} \frac{da}{a^2 H(a)} = \int_0^{a_{\text{dec}}} \frac{da}{H_0\sqrt{\Omega_{r,0}}} = \frac{a_{\text{dec}}}{H_0\sqrt{\Omega_{r,0}}}. \tag{1.57}$$

The comoving distance from here to decoupling, assuming the universe has been mostly matter dominated ($w_m = 0$) throughout this time, is given by

$$\chi_{\text{dec}} = \int_{a_{\text{dec}}}^{a_0} \frac{da}{a^2 H(a)} = \int_{a_{\text{dec}}}^1 \frac{da}{H_0\sqrt{a\ \Omega_{m,0}}} = \frac{2(1-\sqrt{a_{\text{dec}}})}{H_0\sqrt{\Omega_{m,0}}}. \tag{1.58}$$

The angular size of the horizon at decoupling is therefore

$$\theta = \frac{a_{\rm dec}\sqrt{\Omega_{m,0}}}{2\sqrt{\Omega_{r,0}}(1-\sqrt{a_{\rm dec}})} \sim 2^\circ, \tag{1.59}$$

where we used the values of the density parameters in eq. (1.51) and $a_{\rm dec} = 1/(1+z_{\rm dec}) = 1/(1+1090) \sim 10^{-3}$. Given that a spherical surface has about 40000 square degrees, we see that the surface of last scattering is made of more than 10^4 causally disconnected patches of space. If there was not enough time for these regions to communicate before decoupling, how are they so similar?

Solution

Both of the above problems are related to initial conditions: how could the radiation era begin so incredibly flat and homogeneous? We've seen that both issues arise due to the behaviour of the comoving Hubble radius during the early radiation era in the standard big bang cosmology. The fact that $(aH)^{-1}$ is increasing is the root of the puzzles. A simple solution is to conjecture a phase preceding the radiation era in which the comoving Hubble radius was decreasing, $d(aH)^{-1}/dt < 0$. If the factor by which $(aH)^{-1}$ decreased during this phase exceeds the factor by which it has subsequently re-expanded, then we can solve the flatness and horizon problems. Since a decreasing $(aH)^{-1}$ implies that $a'' > 0$ by (1.40), this phase is called *inflation*: it is a brief period of rapid accelerated expansion. The benefit of this period of inflation is that the universe can start from a completely generic initial state, and be driven towards homogeneity and isotropy.

How much inflation do we need? At the least, we require that the observable universe today fits in the comoving Hubble radius at the beginning of inflation: $(a_0H_0)^{-1} < (a_IH_I)^{-1}$. We can relate the present day comoving Hubble radius $(a_0H_0)^{-1}$ to its value at the end of inflation $(a_EH_E)^{-1}$ by simply assuming that the Universe has been radiation dominated since then.

$$\frac{a_0H_0}{a_EH_E} \sim \frac{a_E}{a_0} \sim \frac{T_0}{T_E} \sim \frac{10^{-3}\text{ eV}}{10^{16}\text{ GeV}} \sim 10^{-29}, \tag{1.60}$$

where we assumed the energy scale of inflation was 10^{16} GeV and that the temperature today is $T_0 \sim 2.7$ K $\sim 10^{-3}$ eV. Thus, inflation will be adequate if the comoving Hubble radius shrinks by a factor of 10^{29}: $(a_IH_I)^{-1} > 10^{29}(a_EH_E)^{-1}$. This is typically stated in terms of the number of required "e-folds of inflation":

$$N_{e\text{-folds}} > \log\left(\frac{a_EH_E}{a_IH_I}\right) \sim 67 \tag{1.61}$$

This is the conventional statement that the solution to the horizon problem requires approximately 60 e-folds of inflation.

Slow-roll inflation

So what caused this period of inflation? This is still a wide open question. It's possible, however, that a single scalar field φ with the simplest potential $V(\varphi) = m_\varphi^2\varphi^2/2$ can do the trick [13, 14]. In the toy model picture of inflation that is presented here, this scalar field, the *inflaton*, must have a potential in which it can "roll" monotonically down from a "slow-roll" region to a local minimum with vanishing potential. "Slow-roll" is defined in terms of the slow-roll parameters:

$$\epsilon \equiv \frac{M_P^2}{2}\left[\frac{dV(\varphi)/d\varphi}{V(\varphi)}\right]^2, \qquad \eta \equiv M_P^2\frac{d^2V(\varphi)/d\varphi^2}{V(\varphi)}. \tag{1.62}$$

The slow-roll condition is $\epsilon, \eta \ll 1$, which is satisfied for the simple potential $V(\varphi) = m_\varphi^2\varphi^2/2$ as long as $|\varphi| > M_P$. A homogeneous scalar field has an energy-momentum tensor given by

$$T^\varphi_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\partial_\alpha\varphi\partial_\beta\varphi - V(\varphi)\right]. \tag{1.63}$$

From this expression, recalling the general form of $T_{\mu\nu}$ from eq. (1.26), we can infer the inflaton's energy density and pressure from the time-time and spatial components respectively, yielding,

$$\rho_\varphi = \frac{1}{2}(\varphi')^2 + V(\varphi), \qquad p_\varphi = \frac{1}{2}(\varphi')^2 - V(\varphi). \tag{1.64}$$

We've seen above that a solution to the horizon and flatness problems requires the universe to be temporarily dominated by some fluid with a negative equation of state $w < -1/3$. Hence, this inflaton must obey $p_\varphi < -\rho_\varphi/3$, requiring the potential energy to dominate over the kinetic energy.

We can now examine the evolution of the inflaton in more detail. Substituting p_φ and ρ_φ into the conservation equation (1.35) leads to the equation of motion,

$$\varphi'' + 3H\varphi' + \frac{dV}{d\varphi} = 0. \tag{1.65}$$

This resembles the standard evolution equation for a particle experiencing a potential force $dV/d\varphi$ and a friction force $3H\varphi'$. The inflaton begins in the slow-roll regime: it is strongly overdamped and quickly relaxes to its terminal velocity. In this regime, acceleration φ'' is negligible and the frictional drag approximately balances the potential force,

$$\varphi' \simeq -\frac{1}{3H}\frac{dV}{d\varphi} \tag{1.66}$$

The slow-roll condition $\epsilon \ll 1$ becomes $(\varphi')^2 \ll V(\varphi)$, which implies $w \simeq -1$ and hence $a'' > 0$. The accelerated expansion of the inflationary epoch is manifest as the field gradually rolls down its potential. During this slow-roll phase, we can approximate the spacetime as de Sitter. Eventually, when the field is close enough to the minimum so that $V(\varphi) \sim m_\varphi^2\varphi^2/2$ and

$H < m_\varphi$, the field begins underdamped oscillations $\varphi(t) \propto a^{-3/2} \cos(m_\varphi t)$. In this oscillatory phase, the energy density is decaying as $\rho_\varphi \propto a^{-3}$, implying $w = 0$ and hence $a'' < 0$. In summary, when the inflaton rolls close enough to the minimum where the slow-roll conditions cease to hold, the universe stops accelerating and begins to decelerate.

One last ingredient is needed to ensure that the universe does not end up empty when inflation ends: the inflaton must couple to other forms of matter and energy. The energy stored in the inflaton field should be converted into ordinary particles in the oscillatory decay phase. This is captured in the equation of motion by introducing an inflaton decay rate Γ_φ,

$$\varphi'' + 3H\varphi' + \Gamma_\varphi \varphi' + \frac{dV}{d\varphi} = 0. \tag{1.67}$$

The particles produced by the inflaton decay will mix and interact, eventually reaching thermal equilibrium at a uniform temperature. This process is called *reheating*, and the temperature of this epoch, $T_{\rm reh}$ is determined by the energy density at the end of the reheating process. There exists a wide variety of reheating models that predict different reheating time scales and energy scales. In any case, upon completion of this thermalization procedure, the standard hot big bang cosmology begins.

Generation of inhomogeneities

In addition to offering an explanation to the flatness and horizon problems, inflation also provides an attractive causal mechanism to generate inhomogeneities. The inflationary era amplifies tiny quantum fluctuations and converts them into classical perturbations that act as seeds for the formation of large scale structure in the present universe. In fact, in slow-roll inflation, the same inflaton field is responsible both for the exponential expansion of space, and the generation of inhomogeneities. We are chiefly interested in the spectrum of scalar perturbations (also called curvature perturbations) generated during inflation as these directly couple to matter and radiation and are primarily responsible for the inhomogeneous structure in the universe.

Inflation generates not only scalar perturbations, but also tensor fluctuations in the gravitational metric, termed gravitational waves. These fluctuations are in fact a unique signature of inflation. If we were to observe tensor perturbations in the CMB, it would be the best confirmation of inflation that cosmologists could hope for.

The standard derivation of the scalar and tensor spectra relies on quantizing the perturbations and imposing vacuum initial conditions. Here we simply state the results for the power

spectra produced via slow-roll inflation,

$$\Delta_S^2(k) = \frac{1}{8\pi^2}\frac{H^2}{M_P^2}\frac{1}{\epsilon}\bigg|_{k=aH}, \qquad \Delta_T^2(k) = \frac{2}{\pi^2}\frac{H^2}{M_P^2}\bigg|_{k=aH} \tag{1.68}$$

Both expressions are evaluated at "horizon crossing" defined by $k = aH$. It turns out that both scalar and tensor power spectra are roughly scale invariant, meaning that the above expressions are independent of k. It is conventional to parameterize the deviation from scale invariance with a reference scale k_* and spectral indices n_s and n_t as follows,

$$\Delta_S^2(k) = \mathcal{A}_S\left(\frac{k}{k_*}\right)^{n_s-1}, \qquad \Delta_T^2(k) = \mathcal{A}_T\left(\frac{k}{k_*}\right)^{n_t}. \tag{1.69}$$

In terms of the slow roll parameters, the spectral indices are given by $n_s = 1 - 2\epsilon - \eta$ and $n_t = -2\epsilon$. The tensor-to-scalar ratio is defined as

$$r = \frac{\mathcal{A}_T}{\mathcal{A}_S} = 16\epsilon \tag{1.70}$$

The temperature fluctuations in the CMB are primarily sourced by scalar fluctuations. The Planck satellite [8] has measured the scalar amplitude and spectral index at $k_* = 0.05$ Mpc^{-1} to be

$$\mathcal{A}_S = (2.142 \pm 0.049) \times 10^{-9}, \qquad n_s = 0.9655 \pm 0.0062. \tag{1.71}$$

Tensor fluctuations produced during inflation should also be present in the CMB, but have not yet been measured. A primary goal of current efforts in observational cosmology is to detect the tensor fluctuations produced during inflation in the CMB. This will be discussed further in section 1.6.

In connection with the upcoming chapters, the tensor power spectrum in eq. (1.68) will be used in section 4.5 to determine the amplitude of tensor perturbations produced in bigravity. The scale invariant scalar power spectrum (1.69) and its measured parameters (1.71) will be important for chapter 5 in which we often refer to the power spectrum of the primordial gravitational potential Ψ_i, defined by $\langle\Psi_i(\mathbf{k})\Psi_i(\mathbf{k}')\rangle = (2\pi)^3 P_\Psi(k)\delta^{(3)}(\mathbf{k}-\mathbf{k}')$. This quantity is related to the power spectrum of the curvature perturbation through $P_\Psi(k) = (2/3)^2 2\pi^2\Delta_S^2(k)/k^3$ [7], valid in the radiation dominated era.

Summary

Let us summarize the main features of inflation. Firstly, inflation incorporates a decreasing comoving Hubble radius, providing a causal mechanism for modes to come into contact. During the required ~ 60 e-folds of rapidly shrinking horizon, the universe undergoes explosive expansion, producing a flat and homogeneous universe. Secondly, in the reheating phase, we

have a mechanism for ending inflation and transitioning to the radiation dominated era. Lastly, a striking feature of inflation is that it provides a mechanism to generate variations around the smooth background through quantum fluctuations of the inflaton. The evolution of the perturbations produced during inflation can then be treated with *cosmological perturbation theory*, presented in the next section 1.3.3. Without inflation, the "initial conditions" for the classical perturbations have to be put in by hand. With inflation and reheating, we have a precise prescription for these initial conditions.

Before concluding this section on inflation, we note that there are countless inflationary models that employ more sophisticated potentials, multiple fields, and other complexities [15, 16]. We await more observational evidence to help us discriminate between the various species of inflationary models. Nevertheless, the general paradigm of inflation is accepted by the majority of cosmologists. On the other hand, there are serious criticisms of inflation, which have led to the pursuit of alternative theories that dynamically explain the initial conditions [17, 18, 19, 20, 21, 22]. There is a lot of work to be done on both the theoretical and observation side to construct and constrain viable models of the early universe. Hopefully, there is enough information that we can gather from within our observable horizon to construct a coherent story of the early universe.

1.3.3 Cosmological perturbation theory

GR's highly non-linear nature makes it difficult to deal with unless there is a high degree of symmetry. Fortunately for cosmologists, the observed homogeneity and isotropy of the universe makes the FRW metric adequate for numerous purposes. However, it's true that our universe is not completely homogeneous and isotropic: it is full of structure! It will be nearly impossible to incorporate inhomogeneities and anisotropies in full generality with GR. The best way is to use perturbation methods.

Cosmological perturbation theory is ubiquitously used to study the evolution of inhomogeneities. It provides a map from the initial perturbations produced during inflation, to the anisotropies measured in the CMB, and further to the large scale structure that we see today. Perturbation theory is valid in cosmology because we know that the perturbations are tiny over a wide range of scales. We'll see in section 1.6 that inhomogeneities in the CMB are one part in 10^5, and so must have been even smaller prior to decoupling. Apart from the smallest scales that undergo gravitational collapse to form galaxies, inhomogeneities are small on large scales ($\gtrsim 10$ Mpc), justifying the application of perturbation theory.

Here we simply provide an overview to capture the essence of linear cosmological perturba-

tion theory. The goal is to understand the equations that govern the evolution of perturbations. Let's begin with the perturbed Einstein equation: $\delta G_{\mu\nu} = (8\pi G_N)\delta T_{\mu\nu}$. Under the assumption that all constituents in the universe can be modeled as perfect fluids (see eq (1.26)), and that these fluids do not contain any anisotropic stress (valid for non-relativistic species), the perturbed energy-momentum tensor associated with these sources is,

$$\delta T^0_0 = \delta\rho, \quad \delta T^0_i = (\rho^{(0)} + p^{(0)})v_i, \quad \delta T^i_j = -\delta p \delta^i_j, \tag{1.72}$$

where $\delta\rho$ and δp are the perturbations in energy density and pressure respectively, and $v^i = dx^i/d\tau$ is the coordinate velocity defined as the spatial perturbation of the four-velocity $\delta u^i \equiv v^i/a$. The background quantities, represented with superscripts (0), are assumed to be smooth and time-dependent. Keep in mind that there are several contributions to the energy-momentum tensor such as photons, baryons, neutrinos, and dark matter which we have not distinguished yet.

Let's now consider the dominant contributions to relativistic and non-relativistic matter. For the latter, this is *dark matter* (see section 1.4.1). It is common to introduce the dark matter density contrast $\delta(\mathbf{x}, \tau)$ as

$$\delta(\mathbf{x}, \tau) \equiv \delta\rho_{\rm dm}(\mathbf{x}, \tau)/\rho^{(0)}_{\rm dm}(\tau), \tag{1.73}$$

which is the quantity that cosmologists are typically interested in. Linear perturbation theory is valid when $\delta(\mathbf{x}, \tau) \ll 1$, and a perturbation is termed non-linear when $\delta(\mathbf{x}, \tau) \sim \mathcal{O}(1)$. When deriving the equations, this is employed by neglecting all higher powers of δ.

The time evolution of the density fluctuations are completely decoupled on different scales in linear perturbation theory. This useful property means that each Fourier mode, $\delta(\mathbf{k}, t)$, defined by

$$\delta(\mathbf{x}, \tau) = \int \frac{d^3k}{(2\pi)^3}\delta(\mathbf{k}, \tau)e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{1.74}$$

evolves independently. Assuming statistical homogeneity and Gaussianity of the density field, the density field is described entirely by the *matter power spectrum* $P_\delta(k, \tau)$,

$$\langle\delta(\mathbf{k}, \tau)\delta(\mathbf{k}, \tau)\rangle = \delta(\mathbf{k} - \mathbf{k}')P_\delta(k, \tau). \tag{1.75}$$

This quantity will be especially useful in chapter 5.

For relativistic matter, it is sufficient for our purposes to consider only the photon distribution, f, and its perturbation Θ defined by,

$$f(\hat{\mathbf{p}}, p, \mathbf{x}, \tau) = \frac{1}{\exp\left[\frac{p}{T(\tau)[1+\Theta(\hat{\mathbf{p}},\mathbf{x},\tau)]}\right] - 1}, \tag{1.76}$$

where T is the time-dependent background temperature. Note that the perturbation Θ allows for both inhomogeneities and anisotropies as it depends on both $\mathbf{x}$ and $\hat{\mathbf{p}}$ where $\mathbf{p}$ is the momentum.

With an understanding of the matter perturbations, let's now turn to perturbations of the metric which is more technically involved. One complexity is the gauge (coordinate) dependence. Even though GR is covariant (independent of coordinate choice), splitting the metric into a background piece and a perturbation piece, $g_{\mu\nu} = g^{(0)}_{\mu\nu} + \delta g_{\mu\nu}$, is not a covariant process. There is no uniquely preferred frame of reference in the presence of perturbations. Infinitely many coordinate choices exist for which the metric and the coordinates reduce to their standard FRW form as the perturbations go to zero.

With the issue of covariance in mind, one can either work with gauge invariant quantities that define the perturbations, or simply choose a particular gauge and work in these specific coordinates throughout. In this work, we adopt a specific gauge: the conformal Newtonian gauge (also called the longitudinal gauge). As introduced in equation (1.24), this gauge consists of just one independent function $\Psi(\mathbf{x}, t)$ that represents the perturbation of the gravitational potential.[2] As for the density contrast, perturbation theory is valid as long as $\Psi \ll 1$.

In an effort to be brief, we will skip the derivation of the perturbation equations, referring to the reader to textbooks such as [7, 23]. Rather than give the most general equations, we provide a simple set that is useful for our universe in terms of only dark matter and photons. We are assuming no anisotropic stress and neglecting baryons. Not all components of the Einstein equations are needed to form an independent set of perturbation equations. In Fourier space, the time-time component is sufficient and is given by

$$k^2\Psi + 3\mathcal{H}(\dot{\Psi} + \mathcal{H}\Psi) = -4\pi G_N \left(\frac{\Omega_{m,0}}{a}\delta + 4\frac{\Omega_{r,0}}{a^2}\Theta_0 \right), \tag{1.77}$$

where Θ_0 is the monopole perturbation of the photon distribution, $\Theta_0(\mathbf{x}, \tau) = \frac{1}{4\pi}\int d^2\hat{\mathbf{p}}\, \Theta(\hat{\mathbf{p}}, \mathbf{x}, \tau)$. Further, from the conservation of the energy-momentum tensor (1.14), the relativistic generalizations of the continuity and Euler equations can be derived,

$$\dot{\delta} + ikv - 3\dot{\Psi} = 0, \tag{1.78}$$

$$\dot{v} + \mathcal{H}v + ik\Psi = 0. \tag{1.79}$$

where v is defined by $\partial_i v = v_i$ and is the spatial perturbation of the dark matter velocity (recall eq. (1.72)). Lastly, there is the Boltzmann equation for photons,

$$\dot{\Theta} + ik\mu\Theta - \dot{\Psi} + ik\mu\Psi = 0. \tag{1.80}$$

[2] This gauge actually contains two independent functions, Ψ and Φ, but under the assumption of zero anisotropic stress, we have $\Psi = \Phi$.

where $\mu = \hat{\mathbf{p}} \cdot \hat{\mathbf{k}}$ is the angle between $\mathbf{k}$ and the direction of the photon momentum. By setting appropriate initial conditions (inflation provides these), equations (1.77)-(1.80) can be numerically solved for the metric, dark matter, and photon perturbations, captured by the variables Ψ, δ, v_i, and Θ.

1.4 The dark side of the universe

1.4.1 Dark matter

We have seen that all of the constituents in the universe are modeled in cosmology as perfect fluids (see eq. (1.26)) with different equations of states, focusing so far on matter and radiation (see eq. (1.38)). Let's solely consider the matter sector. All of the non-relativistic particles in stars, planets, and gas can be modeled in this sector. Nearly all matter that is encountered in everyday life is ordinary baryonic matter (made from elements in the period table). A major mystery of modern physics is that the amount of baryonic matter in the universe is not nearly enough to account for our observations. All of the matter that we can see or detect with or telescopes is not enough to explain the formation and dynamics of structure in galaxies and galaxy clusters. There is a mismatch between the matter that we can see and the matter inferred from gravity. Astronomers hypothesized that this mismatch is due to *dark matter*.

Dark matter gets its name because it does not emit or interact with electromagnetic radiation, and thus we can only observe it indirectly through its gravitational effects. From the first proposal for dark matter by Fritz Zwicky in the 1930's based on cluster observations, there is now overwhelming evidence for its existence. This evidence exists at all gravitationally relevant scales, from galaxies to our full horizon. Notable examples are galaxy rotation speeds, velocities of galaxies in clusters, merging galaxy clusters, large scale structure, and the CMB [24].

Given that dark matter seems to be required by a highly interconnected web of observations, it is a crucial component of the current standard cosmological model. The standard model employs *cold* dark matter (CDM) for which the particle speed is non-relativistic. Remarkably, cold dark matter contributes to the universe's energy density budget about 5 times more than baryonic matter. By comparing the relative heights of observed peaks in the CMB temperature power spectrum (see figure 1.3 in section 1.6), the baryon and CDM density parameters, Ω_b, Ω_c, can be obtained. The current results from Planck [8] suggest values of

$$\Omega_b h^2 = 0.02222 \pm 0.00023, \qquad \Omega_c h^2 = 0.1197 \pm 0.0022, \quad \rightarrow \quad \frac{\Omega_c}{\Omega_b} \sim 5.4, \tag{1.81}$$

where h is a dimensionless constant related to the Hubble constant $H_0 = 100h$ km/s/Mpc.

Although it is well established that dark matter exists, there is no consensus on what exactly dark matter is, despite decades worth of effort by theorists and experimentalists. A slew of potential candidates for CDM have been proposed. Particularly appealing categories are axions [25, 26] and WIMPS [27] (Weakly Interactive Massive Particles), although no particles of either type have been detected. However, current direct and indirect detection methods are now entering a sensitivity regime in which some theoretical candidates could be detected [28, 29]. Scientists are optimistic that the next generation of dark matter detectors will yield the next great discovery in modern physics.

Dark matter is a puzzling part of the cosmological model, but it will not be the focus of this work. Instead, we turn to an even bigger source of mystery affecting the largest scales of our universe: dark energy.

1.4.2 Dark energy

The end of section 1.3 motivated the need for a mysterious energy density component. Since CMB obsevations strongly suggest $\Omega_K \sim 0$ [8], dark energy is needed to balance the Friedmann equation (1.49). Different sources of evidence are all consistent. To fit the data, this dark energy must have a negative pressure $w < -1/3$, and it must account for a huge chunk of the universe's energy density $\Omega_{\rm DE} \sim 0.7$. GR allows us to link the geometry and dynamics of spacetime to the matter and energy density of the universe. Here, the link implies that dark energy is responsible for the recent accelerated expansion of the universe $a'' > 0$. So, what is this dark energy?

The simplest model of dark energy is a cosmological constant, denoted by Λ.[3] Dark energy in the form of a cosmological constant Λ, together with cold dark matter (CDM), comprise the basis of the standard cosmological model, "ΛCDM." A cosmological constant means that the dark energy has a constant energy density that fills space homogeneously and does not dilute with the expansion of the universe. It is synonymous with vacuum energy. As a practical explanation, this works quite well, and is consistent with many of the latest observational results. Let's explore how this fits in with the cosmological model described in the previous section.

A cosmological constant is constant in the sense that it does not dilute with expansion like matter and radiation do. The energy density relation analogous to eq. (1.38) is simply

$$\rho_\Lambda = \rho_{\Lambda,0} = \frac{\Lambda}{8\pi G_N}. \tag{1.82}$$

[3] Λ is a dimensionful parameter with units of $[\text{mass}]^2 = [\text{energy}]^2 = [\text{length}]^{-2}$.

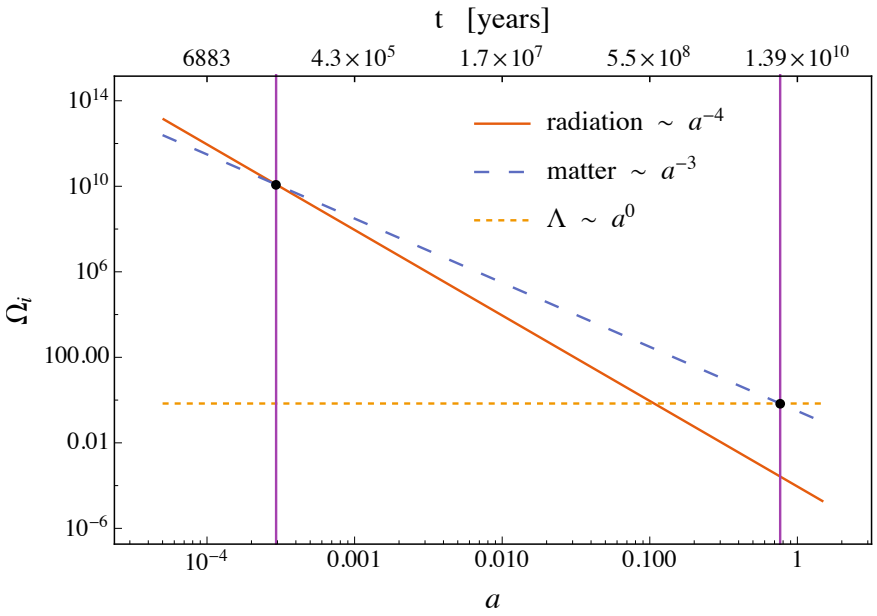

Figure 1.1: The evolution of the density parameter for matter, radiation, and a cosmological constant as a function of scale factor, a, on the lower axis and time, t, since the big bang on the upper axis. The early universe was radiation dominated until $a = \Omega_{r,0}/\Omega_{m,0} \sim 3 \times 10^{-4}$, $t \sim 5 \times 10^4$ years. Matter has dominated for a large portion of cosmic time, until recently, $a = (\Omega_{m,0}/\Omega_{\Lambda,0})^{1/3} \sim 0.76$, $t \sim 10^{10}$ years. Our universe is now dark energy dominated, and has thus begun an epoch of accelerated expansion.

From the general expression in equation (1.36) one can see that the equation of state parameter for the cosmological constant is $w = -1$. The corresponding behaviour of the scale factor with time in a cosmological constant dominated universe is then given in equations (1.45), (1.46). Figure 1.1 shows the evolution of the various density components with scale factor, illustrating the cosmic epochs of radiation, matter, and dark energy domination. The epoch of dark energy domination began "recently" in cosmic history, about 3.6 billion years ago.

Implementing the cosmological constant is a simple fix in Einstein's theory. One interpretation is that the cosmological constant is an additional form of energy which is added to the right hand side of Einstein's equation (1.9),

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G_N T_{\mu\nu} - \Lambda g_{\mu\nu}. \tag{1.83}$$

Here, the cosmological constant acts as a homogenous perfect fluid with equation of state $w = -1$ and corresponding energy momentum tensor,

$$T^{\Lambda}_{\mu\nu} = (\rho^{\Lambda} + p^{\Lambda})u_\mu u_\nu + p^{\Lambda} g_{\mu\nu} \tag{1.84}$$

$$= -\rho^{\Lambda} g_{\mu\nu} \tag{1.85}$$

A different interpretation comes from putting Λ on the left hand side of Einstein's equation,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{1.86}$$

where Λ is a new free parameter of GR. Λ is neither forbidden nor specified by Einstein's original theory, so there is nothing that prevents one from including it in the field equations. As written, (1.86) is considered a modification of gravity since we are directly changing the form of the geometry of spacetime. In the Einstein-Hilbert action (1.16), Λ appears as,

$$S_{\mathrm{GR}+\Lambda} = \int d^4x\sqrt{-g}\left(\frac{M_P^2}{2}R - M_P^2\Lambda + \mathcal{L}_{\mathrm{matter}}\right). \tag{1.87}$$

It is interesting to note that the first use of the cosmological constant term was by Einstein himself. Considering the astronomical data at the time, Einstein hoped his theory would describe a static universe (neither expanding or collapsing). Since a universe filled with only normal matter and described by (1.50) would collapse under gravity, he needed a cosmological constant term to act as a repulsive force to balance it out. When Hubble discovered that the universe is indeed expanding, the desire for a static universe was eliminated. Einstein is famously known for calling this his biggest blunder. Nonetheless, the cosmological constant term remains a legitimate addition to Einstein's original equations. Whether perceived as a modification of spacetime curvature (1.86), or an addition to the matter content of the universe (1.84), the effects of the cosmological constant are the same, and we can simply think of it as a free parameter whose value needs to be constrained by observations.

To describe dark energy and be consistent with all of our observations of the universe (such as the luminosity of supernovae, the observed spatial flatness and structure formation history), we need to set the density parameter of the cosmological constant to be [8],

$$\Omega_\Lambda = 0.6911 \pm 0.0062, \tag{1.88}$$

which gives a cosmological constant value of,

$$\frac{\Lambda_{\mathrm{observed}}}{M_P^2} = \frac{\rho_\Lambda}{M_P^4} = \frac{\rho_{\mathrm{crit}}\Omega_\Lambda}{M_P^4} \simeq 7\times 10^{-121}, \tag{1.89}$$

where we recalled the definition of ρ_{crit} from eq. (1.48) and used the standard conversions in the conventional units ($\mathrm{g/cm}^3 = 8.3\times 10^{90} M_P^4$).

A cosmological constant is the simplest way to explain dark energy. No introduction of complicated dynamics or severe modifications to general relativity are in use. A constant energy density that homogeneously fills empty space is the least invasive way to explain the accelerated expansion of the universe. This abides by a basic tenet of science: adopt the simplest interpretation of the data, and add complications only if forced to by further observation. Further, the predictions made by GR when a tuned cosmological constant is added actually fit most of our observational data extremely well. From the observational point of view, there is no need to consider more complicated forms of dark energy.

Nonetheless, it is natural for cosmologists to consider extensions to the cosmological constant, at least for the sake of having alternatives to test the theory against. The most basic extensions can be considered in terms of the equation of state parameter, w, alone. Although the latest measurements from Planck indicate that w is consistent with a cosmological constant [8], $w = -1.019^{+0.075}_{-0.080}$, we can further allow w to vary with scale factor, and parameterize its Taylor expansion as

$$w(a) = w_0 + (1 - a)w_a. \tag{1.90}$$

The recent estimates for the parameters w_0 and w_a are again consistent with a cosmological constant: $w_0 \sim -1,\ w_a \sim 0$ (see figure 4 in ref [30]). Experimental efforts continue to try and nail this measurement, as even a slight deviation from Λ would be interesting. Notably, the Dark Energy Survey (DES) [31], is currently operating with the main goal of characterizing dark energy by determining the equation of state with more precision. As it currently stands, the cosmological constant appears to be a sensible description of dark energy.

The cosmological constant problem

Despite its practical success, there is a logical inconsistency from the theoretical standpoint: *the cosmological constant problem* (see [32, 33] for reviews). As discussed above, the value of Λ is not specified by GR. From the field theory point of view, however, we have no choice in what the value of the cosmological constant should be. In quantum field theory, the vacuum, like any physical object, has an energy density. The vacuum is maximally symmetric, implying that its energy-momentum tensor is proportional to the metric,

$$T^{(\mathrm{vac})}_{\mu\nu} = -\rho^{(\mathrm{vac})} g_{\mu\nu}. \tag{1.91}$$

Comparing this to (1.84), the vacuum energy has the same form as a cosmological constant. It is therefore common to use the terms "cosmological constant" and "vacuum energy" interchangeably. Using quantum field theory, we can calculate contributions to $\Lambda^{(\mathrm{vac})} = \rho^{(\mathrm{vac})}/M_P^2$ and compare to the observed value.

Like a harmonic oscillator in the ground state, every mode of every free field contributes a zero-point energy to the vacuum's energy density. This energy arises from virtual particle-antiparticle pairs described by loop diagrams. The contributions from each field depend on the cutoff of the effective field theory that is being used. If we are confident that we can trust both theories (standard model of particle physics + gravity) all the way up to the Planck scale, we would receive a contribution on the order of unity,

$$\Lambda_{\mathrm{QFT}} = \frac{M^4_{\mathrm{cutoff}}}{M_P^2} \qquad \rightarrow \qquad \frac{\Lambda_{\mathrm{QFT}}}{M_P^2} \sim 1 \tag{1.92}$$

This "predicted" value of the vacuum energy density based on quantum field theory is about 120 orders of magnitude larger than the observed value $\Lambda_{\text{QFT}} \sim 10^{120}\Lambda_{\text{observed}}$! In addition to this known contribution, there could also be an unspecified bare cosmological constant. The above result requires a bare cosmological constant to be finely tuned to over 120 decimal places so that it cancels this large contribution, leaving us with the tiny observed value needed for cosmology. It is not logically impossible, but this fine tuning is certainly not theoretically pleasing. Note that field theory may fail earlier, implying a lower cutoff scale in (1.92). Regardless of how much we trust the estimated prediction for Λ_{QFT}, and the validity of the calculation, it is clear that there is a tremendous discrepancy.

This enormous fine tuning is termed the cosmological constant problem. Essentially, we lack a fundamental understanding of where the observed value of the cosmological constant comes from. Is the cosmological constant just an unpleasant fact of our universe, or is there some underlying physics that we do not yet fully comprehend?

There are really two issues at hand:

- **Naturalness:** In physics, the dimensionless physical constants appearing in the theory should take values relatively close to unity. This is more of an aesthetic criteria than a physical one. The idea of naturalness comes from the effective field theory notion that all conceivable terms in the effective action that preserve the symmetries of the theory should appear in the action with natural coefficients [34]. Since the observed cosmological constant is many orders of magnitude smaller than one (in planck units), it violates this notion of naturalness.

- **Technical naturalness:** If a small number does exist, physicists prefer that it be technically natural. This means that the small parameter is stable to quantum corrections. Additionally, when a parameter is technically natural, the theory gains an additional symmetry as the parameter is set to zero. This is the case for the small fermion masses: they do not receive large quantum corrections, so their small values are at least stable, and chiral symmetry appears as they are set to zero. For the cosmological constant, however, no known symmetry appears by setting it to zero. It is expected to receive large quantum corrections as in eq. (1.92). The cosmological constant is both unnatural and technically unnatural.

So far we have been discussing what is known as the "old" cosmological constant problem, simply stated as: why is the cosmological constant not large? There's a prediction, and it's big. There are currently no accepted solutions to this problem, and it remains an open area of research in theoretical physics. There is also the slightly less severe "new" cosmological

constant problem: why is the cosmological constant the value we observe it to be, and not zero? There is also the related but separate *coincidence* problem. Referring to figure 1.1, we happen to live in a brief cosmological era in which both matter and vacuum are of comparable magnitudes. This is another unnatural fact about the cosmological constant for which we would like some deeper explanation: why now?

Often, the anthropic principle is employed to explain the cosmological constant problem [35]. The idea is essentially that observers like us will only experience conditions which allow for observers to exist. This assumes that some parameters are not determined by the theory but can actually take on a range of values. We live in a universe in which the particular values are compatible with forming life as we know it. Well before the discovery of accelerated expansion, Weinberg argued in [35] that the cosmological constant should be small based on the anthropic principle. If it were any bigger or smaller, we would not even exist! If its magnitude was too small, the universe would have collapsed under its own gravity before life could form. If its magnitude was too big, the matter in the universe would be ripped apart too fast to form galaxies, stars and planets. Whether or not this explanation is satisfactory is hotly debated in the field.

It is not surprising that the cosmological constant problem has led physicists to pursue alternative explanations for cosmic acceleration. More complicated models for dark energy have been proposed, many of which simply add a dynamic component to the cosmological constant [36], upgrading it to a scalar quintessence field. The fine tuning problem remains unavoidable in any such model.

An alternative approach is to consider more fundamental modifications to the gravitational sector. GR has been successful in explaining and predicting observations on solar system scales, such as orbital precession, gravitational lensing, gravitational redshift, and most recently with LIGO [37], gravitational waves. However, we have seen that employing GR on cosmological scales requires dark matter and dark energy to preserve its success. This may be a signal that we need to go beyond Einstein's theory. An interesting possibility is that all of the evidence for dark energy is actually evidence for a breakdown of GR on cosmological scales.

1.5 Modified Gravity

The unpleasant issues arising within GR on cosmological scales are cause for speculation that GR needs some fundamental modification. Perhaps something is missing in the infrared picture of GR. The idea of modifying gravity on cosmological scales has been an active area of research over the past decade. A vast range of infrared modified theories exist in the current

literature, often termed the modified gravity "zoo" (see [38, 39] for reviews). Some of these include extra scalar, vector, or tensor fields in the gravitational sector, some include higher spacetime dimensions, and some alter the simple Einstein-Hilbert action (1.15) by considering more general actions $R \to f(R)$. There are other classes of alternatives arising from different motivations, such as attempting to quantize gravity, unify gravity with the other fundamental forces, or explain dark matter through modified Newtonian dynamics.

In addition to these motivations for modifying gravity, is the simple desire to learn more about the intricacies of GR. Deforming GR is a great way to discover new structures which could have unexpected applications. Attempts to modify GR so far have uncovered that it is an incredibly simple, rigid, and robust theory. Modifications introduce complications, and while these complications may be able to fix one problem, they do not go without penalty. With even the slightest tweak, it is hard to avoid the appearance of unwanted pathologies.

There are numerous tests that the modified theory must pass to be deemed a viable candidate. From a theoretical standpoint, it should be mathematically consistent and possess well posed equations of motion. It should also be continuous in its parameters: if a theory deforms GR by a small parameter, it should match GR in the limit that the parameter goes to zero. Once this is established, its predictions must match all astrophysical and cosmological data. Ideally, the theory should also be observationally testable against GR and other candidates. Lastly, the theory should be well-motivated. An uninspired fudge factor, such as the cosmological constant, is not as convincing as a modification based on fundamental principles.

One of the most well motivated modifications comes from stripping GR down to its underlying principles from a field theory perspective. Forgetting about the spacetime geometry picture envisioned by Einstein, at its core, GR is the unique theory of a non-trivially interacting massless spin-2 particle, the graviton. The graviton is the particle carrier for the gravitational force, analogous to the spin-0 photon for the electromagnetic force. One of the simplest modifications to consider from here is a theory that propagates a *massive* spin-2 particle. Can the graviton have a non-zero mass? This seems like a natural extension given that we already know the particle carriers of the electroweak forces acquire a mass via the Higgs mechanism.

Massive gravity is a modified gravity theory in which a small non-zero mass, m, is given to the graviton. Like any other modification to GR, massive gravity changes the degrees of freedom. The massless spin-2 graviton of GR propagates 2 degrees of freedom. Turning on the graviton mass means that massive gravity propagates $2s + 1 = 5$ degrees of freedom. With extra degrees of freedom, massive gravity can produce new and exciting features, which have sparked renewed interest in this theory over the past 5 years (for reviews, see [40, 41]).

The main motivation for studying massive gravity is the possibility to explain cosmic acceleration: the extra degrees of freedom can play the role of dark energy. Choosing the graviton mass to be on the order of the Hubble scale, $m \sim H_0 \sim 1.4 \times 10^{-33}$ eV, the theory admits "self-accelerating" solutions without ever explicitly adding a cosmological constant, $\Lambda = 0$. In this case, the acceleration is a fundamental built-in feature of the theory, not an additional factor. It is also interesting to study purely from the theoretical field theory perspective: is it possible to construct a consistent theory of an interacting massive spin 2 particle?

Many attempts to explain cosmic acceleration with modified gravity simply shift the fine tuning problem of the cosmological constant to other parameters of the theory. For massive gravity, the small parameter is the graviton mass, m, itself. However, massive gravity theories have the advantage of possessing a technically natural explanation. The graviton mass, unlike Λ, is a stable parameter that does not receive large quantum corrections. This is consistent with the fact that a symmetry is recovered as $m \to 0$: in the massless limit, the gauge-invariance of GR is restored.

The intuition that a graviton with a Hubble scale mass can modify gravity in the infrared comes from the simple Newtonian gravitational potential, U, describing the gravitational interaction between massive bodies M_1 and M_2. When the force mediating particle is massive, there is an extra Yukawa suppression:

$$U = -\frac{G_N M_1 M_2}{r} e^{-mr}. \tag{1.93}$$

One can see that for $m \sim H_0$, the suppression would cause a weakening of the gravitational force on cosmological scales H_0^{-1}. Meanwhile, on small scales for $r \ll m^{-1}$, we have agreement with the standard potential.

The following chapter 2 will outline the development of massive gravity, from the first attempts by Fierz and Pauli in 1939 [42], to the construction of a fully consistent theory by de Rham, Gabadadze, and Tolley in 2010 [43, 44]. The development was plagued by several peculiarities and obstacles, which will be highlighted. Before delving into the theoretical details which will be presented in chapter 2, we mention that there are experimental limits on the mass of the graviton from the effects of the Yukawa potential, modified dispersion relation, and fifth force, reviewed in [45]:

$$m \lesssim 7.2 \times 10^{-23} \text{ eV}, \qquad \text{precession of Mercury [46, 47]} \tag{1.94}$$

$$m \lesssim 1.2 \times 10^{-22} \text{ eV}, \qquad \text{BH-BH merger GW150916 [48]} \tag{1.95}$$

$$m \lesssim 10^{-32} \text{ eV}, \qquad \text{precession in galileon and DGP theories [49, 50, 51]} \tag{1.96}$$

The lowest bound, although model dependent, is only about one order of magnitude above the present Hubble scale (1.41), which is the value needed to explain accelerated expansion.

1.6 Cosmic Microwave Background

In our efforts to reveal the mysteries of the universe on the largest scales, we must look out as far as we can see, to the edges of our horizon. From here on Earth, the observable edge of the universe is at the surface of last scattering. This "surface" is a prediction of big bang cosmology, and is well understood from basic physics. The very early universe was a hot, dense, fully ionized, uniform plasma. Upon expansion, the universe cooled, and eventually reached the critical *recombination* epoch in which protons and electrons combined to form neutral hydrogen. *Photon decoupling* occurred shortly after, in which photons could freely stream through space rather than be constantly scattered by ions in the plasma. During this era, the universe went from an opaque plasma fog to transparent.

At the time of decoupling, the universe was about 379000 years old, about 3000 K hot, and about one thousandth of its current size ($a_{\text{dec}} \sim 10^{-3}$), corresponding to a redshift of $z_{\text{dec}} = 1090$ (see eq (1.34)). After decoupling, the photons have been propagating freely through space ever since. These photons are the source of the relic radiation from the early universe, conventionally termed the cosmic microwave background (CMB) radiation. This radiation is observed by us to come from a spherical shell called the last scattering surface.

As the universe continually expanded, the photons redshifted and cooled, decreasing in energy and increasing in wavelength as in equation (1.31). The CMB radiation is detected by us on Earth with a uniform temperature $T_0 = 2.72548 \pm 0.00057$ K [52]. The measured frequency profile indicates thermal equilibrium and it is the best example of a perfect blackbody in the universe. The first discovery of the CMB radiation by radio astronomers Arno Penzias and Robert Wilson in the 1960's was awarded the Nobel Prize [53].

The CMB is the oldest light in the universe, and so it carries a wealth of information about the primordial universe. The radiation we observe today is a snapshot of the local properties of the gas of CMB photons at the time of decoupling, such as density, peculiar velocity and the total gravitational potential. These quantities are related to the primordial perturbations by the simple acoustic physics describing the pre-recombination plasma. As observers, we can measure the radiation's frequency spectrum, temperature, and polarization. Each of these observables are fully packed with cosmological information. By mapping the CMB across the sky, cosmologists thus hope to answer some of the biggest questions in physics.

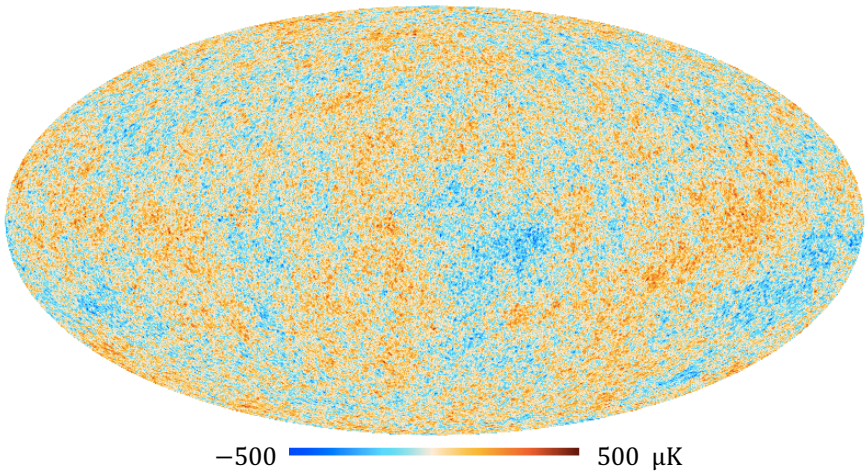

Figure 1.2: Full sky map of the fluctuations in the CMB temperature, determined by the Planck satellite [56]. The temperature is fairly uniform, averaging $T_0 \sim 2.73$, with tiny fluctuations on the order of 10^{-4} K. The hot and cold spots trace the over and under densities generated by primordial quantum perturbations. The over densities have evolved and collapsed under gravity to form the dense structure in the universe that we see today.

Temperature Anisotropies

The vast majority of information lies in the CMB temperature field. The average temperature is amazingly uniform across the sky, however, sensitive probes have revealed small fluctuations on the order of 10^{-4}. These hot and cold spots signify the over and under densities produced by quantum fluctuations in the early universe. The over densities are the seeds that evolved under the gravitational instability of cold dark matter to form the large-scale structure in the universe.

Large-angle temperature anisotropies were detected first by COBE [54] in 1992, followed by an impressive leap in precision by WMAP [55] in the next decade. The current state of the art dataset is provided by the Planck satellite [4], and contains a high resolution full-sky map of these temperature fluctuations. We generally define a dimensionless quantity $\Theta(\hat{\mathbf{n}}) = (T(\hat{\mathbf{n}}) - \langle T \rangle)/\langle T \rangle$ to capture deviations from the average temperature $\langle T \rangle$ in a given direction on the sky, $\hat{\mathbf{n}} = (\theta, \phi)$.

As illustrated in figure 1.2, we observe these temperature fluctuations projected on a 2D spherical sky surface, and so it is convenient to expand the temperature field in spherical harmonics defined by,

$$Y_{\ell m} = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta) e^{im\phi}, \tag{1.97}$$

where the ℓ runs from 0 to infinity, $-\ell \leq m \leq \ell$, and P_ℓ^m are the associated Legendre functions. The multipole ℓ represents a given angular scale on the sky of π/ℓ degrees. Expanding the

temperature anisotropy using these functions gives,

$$\Theta(\hat{\mathbf{n}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}), \quad \text{where} \quad a_{\ell m} = \int_{\theta=-\pi}^{\pi} \int_{\phi=0}^{2\pi} d^2\hat{\mathbf{n}}\ \Theta(\hat{\mathbf{n}}) Y^*_{\ell m}(\hat{\mathbf{n}}) \tag{1.98}$$

Note that the monopole ($\ell = 0$) which is the average temperature over the whole sky, and dipole ($\ell = 1$) which depends linearly on the velocity of the observer, are typically removed from the sum when analyzing data. The upper bound of the sum is dictated by the resolution of the experiment. The expansion coefficients $a_{\ell m}$ are called the multipole or harmonic coefficients.

Theory, which stems from quantum mechanics in the early universe, only allows us to predict the statistical properties of cosmological fields, not their exact form. Our observations represent just one particular realization of a random process. The statistical properties of the temperature fluctuations should respect the symmetries of the background model. In the case of the FRW model introduced in section 1.3.1, this means the statistics should respect homogeneity and isotropy. Demanding invariance under rotations fixes the second-order statistics to be of the form

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell^{TT}, \tag{1.99}$$

which defines the power spectrum, C_ℓ^{TT}, of the temperature fluctuations. The above average is taken over an ensemble of realizations of the fluctuations. The simplest models of inflation predict that the fluctuations should also be Gaussian at early times, which means that the multipole coefficients have Gaussian distributions with zero mean and variance equal to C_ℓ^{TT}. In this case, all we need to characterize the statistics of the temperature anisotropies is the power spectrum (higher point statistics will be zero). Efforts to date have revealed no evidence for primordial non-Gaussianity [57]. Therefore, measuring the anisotropy power spectrum has been a main goal of CMB research.

Figure 1.3 displays the latest power spectrum measurements from Planck. It is conventionally plotted as $\ell(\ell+1)C_\ell^{TT}/(2\pi)$ so that there is equal power per log interval in ℓ. The peak structure is generated from physical laws that govern the oscillations in the pre-recombination photon-baryon plamsa. In this plasma, gravity tries to compress the fluid in potential wells, while photon pressure resists compression. The resulting oscillations are called sound waves or acoustic oscillations. They cause a spatial variation in the temperature that is imprinted on the CMB.

The position, shape, and height of the peaks in the power spectrum tell us an extraordinary amount about our universe, allowing us to measure the density parameters Ω_i defined in equation (1.47). For instance, the position of the peaks is mainly sensitive to curvature. Based on the position of the first peak, we know that our universe is consistent with being flat (zero curvature). Since the Friedmann equation (1.49) with $\Omega_K = 0$ implies that the total

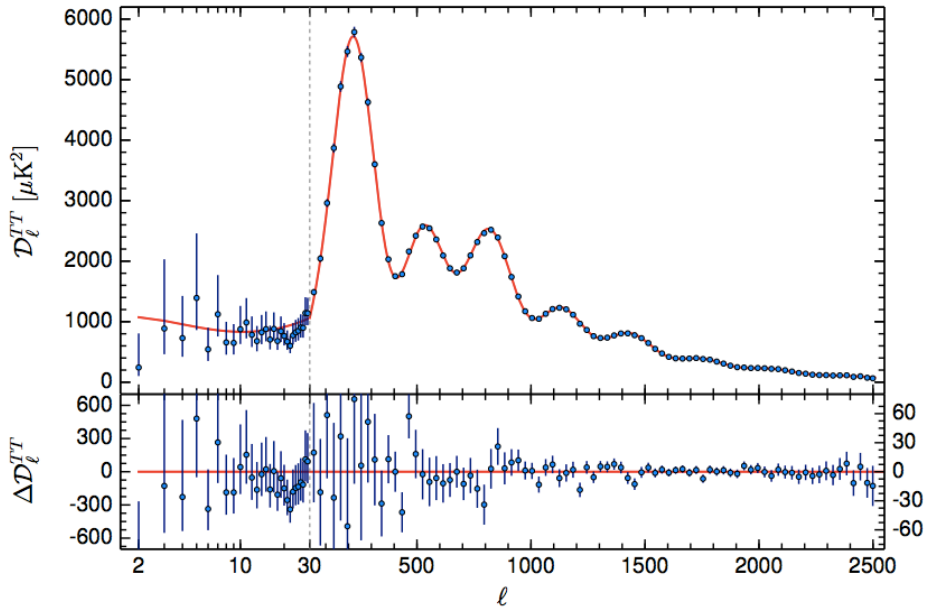

Figure 1.3: Adapted from Planck [8]: The power spectrum $\mathcal{D}_\ell^{TT} \equiv \ell(\ell+1)C_\ell^{TT}/(2\pi)$ is plotted as a function of multipole moment, related to the angular scale as $\theta \sim \pi/\ell$. The dots are measurements made by Planck, with error bars representing $\pm 1\sigma$ uncertainties. The curve in the upper panel is the best fit of the standard cosmological model ΛCDM, fit with six parameters. The lower panel shows the residuals with respect to this model. Notice the remarkable fit of the theory to the data points.

energy density is close to critical, and many observations indicate that the matter energy density is sub-critical, we see that dark energy is needed to make these statements consistent. Furthermore, the baryon and dark matter densities affect the oscillatory behaviour of the plasma. Raising the baryon density enhances every other peak in the power spectrum, while raising the dark matter density reduces the overall amplitude of the peaks. The shapes and heights of the second and third peak tell us that the density of baryons and dark matter in our universe are the values given in equation (1.81). For more details on the relationships between the power spectrum peaks and cosmological parameters, see references [58, 59, 60].

Cosmic Variance

With noise-free full-sky measurements of the temperature field, the CMB power spectrum can be estimated by,

$$\hat{C}_\ell^{TT} = \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2. \tag{1.100}$$

This is an unbiased estimator, $\langle \hat{C}_\ell^{TT} \rangle = C_\ell^{TT}$, but there is an unavoidable error in the estimation of any given C_ℓ^{TT} of $\Delta C_\ell^{TT} = \sqrt{2/(2\ell+1)}$. This limit on precision is called *cosmic variance*. This is essentially a sampling variance stemming from the fact that we only have one universe

and one location to make measurements from. As mentioned above, our theory does not predict the exact form of any cosmological field. It only makes predictions for the probability distribution from which the one realization that we actually can observe is drawn. Any estimate of properties of the underlying probability distribution will inevitably have some random error, due to our attempt to estimate ensemble-averaged quantities from a single realization. In the above estimate of the power spectrum, we are intrinsically limited on the number of independent m-modes we can measure as there are only $2\ell + 1$ of these for each multipole.

The Planck satellite is mostly limited by cosmic variance out to $\ell \sim 2000$. Its resolution is sufficiently high and the noise is sufficiently low to make the error bars in figure 1.3 dominated by cosmic variance. However, cosmic variance especially hinders our ability to learn about large scale (low multipole ℓ) properties of universe. This is evident looking at the size of the error bars in figure 1.3 for $\ell < 30$.

Polarization

Photons in the pre-recombination plasma were constantly being scattered by ions. This scattering produces polarization of CMB photons, much like how incoming light from the sky is polarized due to scattering in our atmosphere. Polarization is an important observable of the CMB, providing complementary information to the temperature fluctuations. Polarization also helps us disentangle the various mechanisms that produce temperature anisotropies, as different sources produce different specific patterns in polarization. Therefore, although the polarization signal is smaller (about 10 times smaller than temperature), it contains directional information, which provides a way of isolating a possible primordial gravitational wave signal. See [61] for an introduction on CMB polarization. Certainly, the next frontier of CMB research involves precise polarization measurements.

Linear polarization of the CMB is generated by Thomson scattering of the *quadrupole* of the temperature anisotropy at the surface of last scattering. The quadrupole moments are given by a_{2m}. There is another contribution to polarization coming from Thomson scattering during the epoch or reionization, $z \sim 6 - 10$, in which the first stars and galaxies formed. Polarization peaks at angular scales that correspond to the angle subtended by the mean free path at last scattering, since these modes have the largest quadrupole anisotropy at last scattering. The quadrupole could arise from three types of perturbations. Scalar perturbations, due to energy density fluctuations, vector perturbations, due to vorticity, and tensor perturbations, due to gravitational waves, all generate quadrupole anisotropies. As mentioned above, these different sources produce different polarization patterns. The different patterns are separated into

"electric" (E), or gradient, and "magnetic" (B), or curl, components. Scalar perturbations produce E-mode polarization, vector perturbations produce mainly B-modes, and tensor perturbations produce both. Importantly, B-modes cannot be produced solely from scalar perturbations, but are expected to arise from inflation (see section 1.3.2).

The E and B nomenclature reflects the global parity: E-modes have electric parity, $E_{\ell m} \to (-1)^{\ell} E_{\ell m}$, while B-modes have magnetic parity, $B_{\ell m} \to (-1)^{\ell+1} B_{\ell m}$. E-mode polarization is directed along or perpendicular to the direction in which its magnitude varies, whereas for B-modes, it is rotated by 45 degrees. We can analyze the E and B maps by decomposing them in terms of spherical harmonics as we did for the temperature anisotropies in equation (1.98), and define power spectra as in equation (1.101),

$$\begin{aligned} \langle E_{\ell m} E^*_{\ell' m'} \rangle &= \delta_{\ell\ell'} \delta_{mm'} C_\ell^{EE}, \\ \langle B_{\ell m} B^*_{\ell' m'} \rangle &= \delta_{\ell\ell'} \delta_{mm'} C_\ell^{BB}, \\ \langle a_{\ell m} E^*_{\ell' m'} \rangle &= \delta_{\ell\ell'} \delta_{mm'} C_\ell^{TE}. \end{aligned} \tag{1.101}$$

E-modes are correlated with temperature, whereas statistical isotropy and parity invariance imply that there are no expected correlations between B and either Θ or E. The current best measurements of the EE and TE power spectra from Planck [8] show remarkable agreement between the polarization measurements and the predicted theory spectra based on the best-fitting ΛCDM model to the temperature anisotropies. Currently, there are only upper limits on the B-mode power spectrum.

A significant detection of B-mode polarization would have profound implications for cosmology. Since B-modes are not produced by density perturbations, a detection of B-mode polarization would give direct information about primordial gravitational waves, which were presumably produced during an inflationary epoch in the early universe as described in section 1.3.2. As of now, we can put a bound on the ratio of the amplitude of tensor perturbations to scalar perturbations (see eq. (1.69)): $\mathcal{A}_T/\mathcal{A}_S < 0.11$ [8]. Characterizing the B-mode polarization is a primary goal of many current ground-based, balloon, and satellite experiments.

Importance for dark energy and modified gravity

The CMB anisotropies open a window to the primordial universe. In addition to unveiling mysteries of the early universe (such as properties of inflation), it should be stressed that the CMB encodes crucial information about the matter composition and geometry of the universe on the largest scales. Thus, it is certainly relevant for our endeavour of learning about dark

energy and modified gravity. There are a number of important effects that dark energy and infrared modified gravity can have on the CMB.

Firstly, alternative dark energy models will change the expansion history of the universe, and hence, change the distance to the surface of last scattering. This would manifest as a shift in the peaks of the power spectrum. The integrated Sachs-Wolfe effect [62], the decay of gravitational potentials at late times, would also alter the power spectrum at low ℓ. Non constant dark energy or modifications of GR would affect the lensing potential [63], as well as the growth rate of structure, potentially leading to a discrepancy between CMB measurements and late-time large scale structure measurements [64, 65]. Additionally, there could be an influence on CMB polarization, for instance if modified gravity models change the sound speed of gravitational waves [66]. See [30] for details on many possible CMB signals that can allow us to constrain dark energy and modified gravity models.

An abundance of different theoretical models for dark energy and modified gravity have been proposed in the literature. The effort of comparing these models and their vast parameter spaces with data is still in its early stages. Given that the CMB is currently our cleanest probe of large scales, it is helpful in its ability to place tight constraints on alternative models, especially when used in combination with other cosmological probes. Next, we discuss how secondary CMB effects can provide additional information about cosmology in the infrared.

Secondary effects

The anisotropies in temperature and other fields that were generated at the surface of last scattering are called *primary* effects. In addition to these primordial anisotropies arising from quantum fluctuations in the early universe, there are several additional processes that add structure to the CMB after recombination (see [67] for a review). On their journey from the surface of last scattering to Earth, CMB photons interact with cosmic structures, causing their frequency, energy or direction of propagation to be affected. These post-decoupling interactions generate *secondary* anisotropies.

As we've seen above, the primary CMB is cosmic variance limited on large scales. This means that we can not learn any more about very large-scale inhomogeneities from the primary CMB alone. Given that nearly all of the hints we have of departures from ΛCDM are on the very largest scales (for a recent summary of CMB anomalies see [68]), there is strong motivation to go beyond the primary CMB to learn more. Additionally, as introduced in the previous sections, more information on large scales is essential to distinguish between various infrared modifications of GR and models for dark energy. Since we have extracted all possible

large-scale information from the primary CMB, careful modeling of secondary effects is now required for constraining cosmology on large scales. The valuable information contained in these secondary signals has prompted a number of advanced surveys [69, 70, 71, 72, 73, 74, 75].

Secondary effects fall under two main categories: gravitational effects from the interactions of photons with gravitational potential wells, and scattering between photons and free electrons. Referring the reader to [67] and [76] for detailed descriptions of the physics of each effect and their corresponding signatures, we simply list the main contributions here. The integrated Sachs-Wolfe effect [62], the Rees-Sciama effect [77], and gravitational lensing of the CMB [78] fall under the first category. Within the scattering category are variations of the Sunyaev-Zeldovich (SZ) effect [79]. The *thermal* SZ effect due to scattering from hot gas in clusters causes a spectral distortion in the blackbody spectrum. The *kinetic* SZ effect due to scattering from ionized gas that is moving with respect to the CMB rest frame produces secondary temperature anisotropies [80]. Additionally, scattering induces secondary polarization as well.

In chapter 5, we focus on the kinetic Sunyaev-Zeldovich (kSZ) effect, and how it can be a potentially powerful probe of large scales. Excitingly, we have recently entered an era in which this effect is detectable with current experiments, and measurements will continue to improve with future probes. The kSZ effect was first detected in 2012 by looking for the contribution to CMB temperature anisotropies induced by the pairwise motion of clusters [81, 82, 83, 84]. Next-generation "Stage 3" and "Stage 4" CMB experiments [85, 86] will have the ability to make high-significance measurements of the kSZ effect. Realizing the full potential of kSZ measurements will rely heavily on cross correlations with probes of large scale structure, making the dramatic improvements to come with the next generation of redshift surveys and 21cm measurements equally important. Such cross correlations open the door to determining the contribution to the global kSZ signal from different redshifts, a technique known as kSZ tomography [87, 88, 89, 90, 91, 92, 93, 94].

The science case for precision measurements of the kSZ effect is quite broad. In addition to revolutionizing our understanding of reionization, it has the power to probe missing baryons e.g. [87, 95, 96], make precision tests of gravity [97, 98, 99], probe anomalous bulk flows [100, 101, 102, 103, 104, 105, 106], constrain the properties of dark energy and dark matter [107, 108], constrain the masses of neutrinos [109], test the Copernican principle [110, 89, 111], constrain the present day vacuum decay rate [112], and test the hypothesis that we inhabit an eternally inflating multiverse [113].

Chapter 5 explores a novel large-scale regime of the kSZ effect. We highlight the ability of kSZ tomography to overcome cosmic variance, allowing us to probe the large scale homogeneity of the universe and test deviations from ΛCDM.

Chapter 2

Massive gravity history and formalism

Theoretical advancements have led to a recent resurgence of interest in massive gravity, an infrared modification to GR in which the graviton has a non-zero mass m. Following the introduction of massive gravity in section 1.5, this chapter describes the historical development of the theory, noting how several setbacks along the way were overcome. We highlight many important features of massive gravity. For instance, section 2.2.2 focuses on the decoupling limit of massive gravity, which allows us to decouple the 5 degrees of freedom of the massive graviton and make a more direct comparison to GR. In this limit, the theory resembles GR modified by an additional scalar *galileon* field. General properties of galileons will be discussed in section 2.2.3. This chapter also describes notable theories with close ties to massive gravity, the DGP Model in 2.3, as well as bigravity 2.4. These theories, and other aspects discussed in this foundational chapter, will be important for the work that follows in chapters 3 and 4 respectively. For recent detailed reviews on massive gravity, see Refs. [40, 41].

2.1 From linear to non-linear massive gravity

Linear Fierz-Pauli gravity

The story of massive gravity begins with Fierz and Pauli, who constructed a linear theory for a massive spin-2 particle in 1939. The Fierz-Pauli action [42] for a single massive spin-2 field in flat-space is given by:

$$S_{\rm FP} = M_P^2 \int d^4x \left[-\partial_\lambda h_{\mu\nu} \partial^\lambda h^{\mu\nu} + \partial_\mu h_{\nu\lambda} \partial^\nu h^{\mu\lambda} - \partial_\mu h^{\mu\nu} \partial_\nu h + \frac{1}{2} \partial_\lambda h \partial^\lambda h \right.$$
$$\left. - \frac{1}{2} m^2 \left(h_{\mu\nu} h^{\mu\nu} - h^2 \right) + \frac{h^{\mu\nu} T_{\mu\nu}}{M_P^2} \right], \tag{2.1}$$

where $h_{\mu\nu}$ is a rank-2 covariant tensor representing the graviton particle, m is the graviton mass, and $T_{\mu\nu}$ is some energy momentum tensor. The indices of $h_{\mu\nu}$ are moved up and down with $\eta_{\mu\nu}$ so that the trace is $h = h_{\mu\nu}\eta^{\mu\nu}$. The first four terms above are simply obtained by expanding the Einstein-Hilbert action (1.15) to quadratic order in $h_{\mu\nu}$ around flat spacetime, ie. taking $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. These terms describe a massless helicity-2 graviton and have the gauge symmetry $\delta h_{\mu\nu} = \partial_\mu \xi_\nu(x) + \partial_\nu \xi_\mu(x)$. Look explicitly now at the Fierz-Pauli mass term which breaks gauge invariance,

$$S_{{\rm FP},m} = -\frac{1}{2} \int d^4x \ m^2 M_P^2 \left(h_{\mu\nu} h^{\mu\nu} - h^2 \right). \tag{2.2}$$

These are the only two possible covariant quadratic terms in $h_{\mu\nu}$. The relative minus sign between the two mass terms is called the Fierz-Pauli tuning. It serves as an additional constraint in order to attain the correct number of degrees of freedom for a massive graviton, 5. Violating this tuning results in an extra unstable degree of freedom [114]. The action then would describe a scalar ghost (with negative kinetic energy) in addition to the massive graviton.

It can be shown [40] that the Fierz-Pauli action indeed propagates 5 degrees of freedom, and thus correctly describes a massive spin-2 graviton. In the massless limit, these 5 degrees of freedom break down into the 2 helicity states of a massless graviton, the 2 helicity states of a massless vector, and 1 single massless scalar. The scalar, which is recognized as the longitudinal mode of the graviton, is still coupled to the trace of the energy-momentum tensor in the massless limit. So taking $m \to 0$ in the Fierz-Pauli action does not quite recover GR, but rather describes a massless graviton plus a coupled scalar. This coupled scalar is responsible for an unacceptable discontinuity in the theory, known as the vDVZ discontinuity [115].

There are various ways to see the vDVZ discontinuity explicitly. For instance, Fierz-Pauli gravity predicts a 25% smaller angle at which light is bent around an object compared

to the GR prediction [40], a deviation much too large to be compatible with current light bending measurements by the sun [116]. The discontinuity appears in the post-Newtonian parameter $\gamma = \Psi/\Phi$ where Ψ and Φ are the metric perturbations away from flat space: $ds^2 = -(1+\Phi)dt^2 + (1+\Psi)dr^2 + \cdots$. In GR $\gamma = 1$, but in linear Fierz-Pauli $\gamma = 1/2$, and since γ goes into the equations for the physical predictions of light bending and other phenomena such as the emission of gravitational radiation, this discontinuity in γ is physically unacceptable.

Although Fierz-Pauli has an incurable discontinuity, it is only a linear theory (the resulting equations of motion are linear). We should expect that a theory of gravity that is able to approach GR in high curvature regimes will be non-linear. Perhaps adding non-linear terms can provide a solution.

Non-linear extensions

A natural starting point for constructing a non-linear massive gravity theory is to use the usual Einstein-Hilbert action (eq. (1.15)) involving the dynamical metric $g_{\mu\nu}$, for the kinetic term and then simply add an appropriate mass term. So what would an appropriate mass term be? One would first imagine that the mass term for $g_{\mu\nu}$ should only depend on $g_{\mu\nu}$ itself in a non derivative way $\sim m(g_{\mu\nu})^2$. However, the only such non trivial term is simply proportional to the volume element $\sqrt{-g}$, and is therefore a cosmological constant. This clearly does not give any mass to the graviton since it does not break gauge invariance. Further, expanding this term to quadratic order around an arbitrary background does not have the Fierz-Pauli form.

The desired mass term requires the introduction of some extra field besides the metric $g_{\mu\nu}$ [117]. One possibility is to add another spin-2 field, $f_{\mu\nu}$, that can be either dynamical or non-dynamical. These two metrics interact through an interaction term in the action, denoted $S_{\rm int}[f,g]$. This interaction term will be the desired mass term for the graviton. For now, we just consider the case where $f_{\mu\nu}$ is a non-dynamical fixed field. In section 2.4, the possibility of adding dynamics to $f_{\mu\nu}$ is explored. All together, following [117] the total action of the theory that is usually considered is:

$$S_{\rm nonlin} = \int d^4x\sqrt{-g}\left(\frac{M_P^2}{2}R(g) + \mathcal{L}_{\rm matter}(g)\right) + S_{\rm int}[f,g]. \tag{2.3}$$

The first term is the usual Einstein-Hilbert action (1.15). The term denoted $\mathcal{L}_{\rm matter}(g)$ is a generic matter Lagrangian assumed to have minimal coupling to $g_{\mu\nu}$ (but not to $f_{\mu\nu}$). The last term is the interaction term which is chosen such that:

1. The theory is generally covariant under diffeomorphisms (common to the two metrics),

2. It has flat space-time as a solution of the field equations for $g_{\mu\nu}$, and

3. It reduces to the Fierz-Pauli form when we take $f_{\mu\nu} = \eta_{\mu\nu}$ and $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, and expand to quadratic order in $h_{\mu\nu}$

In [118] it is shown that simply requiring that f and g interact in a non derivative way and that the theory is diffeomorphism invariant, the necessary form of $S_{\rm int}[f,g]$ is

$$S_{\rm int}[f,g] = m^2 M_P^2 \int d^4x \mathcal{V}(g,f) \equiv m^2 M_P^2 \int d^4x \sqrt{-g} V(g^{-1}f), \tag{2.4}$$

for some suitable "potential" V. We will see more specific forms for $S_{\rm int}[f,g]$ in the DGP model and dRGT model in the following sections.

2.2 Development of dRGT Massive Gravity

In attempting to resolve the vDVZ discontinuity with a non-linear extension of Fierz-Pauli massive gravity, new problems arise. Boulware and Deser [119] discovered that a broad range of fully non-linear massive gravity theories possess ghost-like instabilities. Fierz-Pauli's linear theory had 5 degrees of freedom (the correct number for a massive spin-2 field), but the non-linear theories under study happened to have 6 degrees of freedom.

Boulware and Deser showed this explicitly through a Hamiltonian analysis in the ADM language [120]. Here we will just highlight the main points. In the ADM language, the *lapse* and *shift* are defined respectively as $N \equiv 1/\sqrt{-g^{00}}$ and $N_i \equiv g_{0i}$ in terms of the metric $g_{\mu\nu}$. In GR, the lapse and the shift are the Lagrange multipliers associated with diffeomorphism invariance, and they generate first class constraints which eliminate 4 out of the 6 possible dynamical degrees of freedom of $g_{\mu\nu}$. Thus, just the 2 usual polarizations of the massless graviton remain. However, in non-linear massive gravity, the addition of a mass term breaks diffeomorphism invariance and modifies the nature of the lapse and shift in the Lagrangian. As Boulware and Deser first pointed out in [119] neither N_i nor N appear as Lagrange multipliers in a general non-linear theory, but rather as auxiliary fields. Using their equations of motion, their values can be solved algebraically and plugged back into the action, resulting in an action with no constraints or gauge symmetries at all. So all 6 degrees of freedom are active. The extra mode is the "Boulware-Deser ghost", which manifests as a scalar field with the wrong sign kinetic term.

It was once thought that this ghost is unavoidable in non-linear massive gravity theories [121], but this conclusion was too quick. The fact that generic interactions give rise to an extra degree of freedom does not imply that all interactions have this issue. It has been recently

shown that it is in fact possible to add appropriate interactions that completely eliminate the ghost [44, 43, 122]. This is the case for dRGT gravity, which was built specifically to be ghost-free by de Rham, Gabadadze, and Tolley. The theory was found using modern effective field theory techniques to find appropriate interactions order by order, and then re-summing the result.

2.2.1 The dRGT Action

We have seen that the massive gravity theories so far have some unattractive features, such as discontinuities and ghost instabilities. The question is whether these problems can be eliminated by choosing a specific form of $V(g^{-1}f)$ from equation (2.4). de Rham, Gabadadze and Tolley examined this [44], and constructed a potential in which the Boulware-Deser ghost disappears. As an important bonus of this potential, the cutoff of the theory is raised to $\Lambda_3 = (m^2 M_P)^{1/3}$. This is an improvement from what was previously believed to be the cutoff: $\Lambda_5 = (m^4 M_P)^{1/5}$ [123]. Raising the cutoff means increasing the size of the regime in which we can trust the effective field theory.

To describe the special form of the potential in the dRGT theory, we start by introducing the functions e_k whose argument is an arbitrary $n \times n$ matrix $X^I{}_J$, $I, J \in \{1, \ldots, n\}$:

$$\begin{aligned}
e_0(X) &= 1, \\
e_1(X) &= [X], \\
e_2(X) &= \frac{1}{2}([X]^2 - [X^2]), \\
e_3(X) &= \frac{1}{6}([X]^3 - 3[X][X^2] + 2[X^3]), \\
e_4(X) &= \frac{1}{24}([X]^4 - 6[X]^2[X^2] + 3[X^2]^2 + 8[X][X^3] - 6[X^4]), \\
&\vdots \\
e_k(X) &= \frac{1}{k!} X^{I_1}{}_{[I_1} \ldots X^{I_k}{}_{I_k]}
\end{aligned} \tag{2.5}$$

where $[X]$ is the trace of the matrix $X^I{}_I$, and the brackets around the indices in the last expression indicate the unnormalized antisymmetric sum over permutations. Note that, by the Cayley-Hamilton theorem, $e_n(X) = \det(X)$ for an $n \times n$ matrix.

dRGT theories are defined by an action of the form [122]:

$$S_{\text{dRGT}} = -\frac{M_P^2}{2} \int d^4x \sqrt{-g} R(g) + M_P^2 m^2 \int d^4x \sqrt{-g} \sum_{n=0}^{4} \tilde{\beta}_n e_n \left(\sqrt{g^{-1} f} \right) \tag{2.6}$$

where $\tilde{\beta}_n$ are arbitrary parameters of the theory, and the above square root is a matrix square root of the tensor $g^{-1}f$. Although it may appear that in 4D, there are 5 parameters $\tilde{\beta}_0, \ldots, \tilde{\beta}_4$

in the theory, there are actually just 2. This is because $\tilde{\beta}_0$ describes a cosmological constant (which does not give any mass to the graviton since it does not break general covariance). In addition, the term proportional to $\tilde{\beta}_4$ does not give any contribution to the field equations for $g_{\mu\nu}$ since $\sqrt{-g}e_4(\sqrt{g^{-1}f}) = \sqrt{-g}\det(\sqrt{g^{-1}f}) = \sqrt{-f}$. So there are three free parameters $\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3$, which become just a two parameter family once the mass of the graviton is fixed. This is easier to see using an alternative but equivalent form of the above action. Following [117], the two parameters become manifest by defining the matrix $\mathcal{K} = \mathbb{1} - \sqrt{g^{-1}f}$, and setting the beta parameters to be $\tilde{\beta}_0 = -6 - 4\alpha_3 - \alpha_4$, $\tilde{\beta}_1 = 3 + 3\alpha_3 + \alpha_4$, $\tilde{\beta}_2 = -1 - 2\alpha_3 - \alpha_4$, $\tilde{\beta}_3 = \alpha_3 + \alpha_4$, $\tilde{\beta}_4 = -\alpha_4$. The dRGT action now reads

$$S_{\text{dRGT}} = -\frac{M_P^2}{2}\int d^4x\sqrt{-g}R(g) - M_P^2 m^2 \int d^4x\sqrt{-g}\,[e_2(\mathcal{K}) + \alpha_3 e_3(\mathcal{K}) + \alpha_4 e_4(\mathcal{K})], \quad (2.7)$$

which has two parameters α_3 and α_4.

2.2.2 The Decoupling Limit

The decoupling limit is a limit in which we take $m \to 0$ and decouple the extra degrees of freedom from gravity as much as possible. Care needs to be taken to properly define this limit so that the resulting theory is not trivial; it still should retain the scalar's self-interactions and couplings to matter (we'll soon see where this scalar mode comes from). This section will outline the derivation of the decoupling limit for the dRGT theory, and discuss its significance.

Starting with the action in the form of (2.7), we perform what is known as a 'Goldstone boson expansion, as introduced in [123]. This is the same as the so-called Stueckelberg trick [40]. Since the action being considered has gauge invariance, we can write the flat background metric $f_{\mu\nu}$ in various coordinate systems. For instance, one can use the following expression for f:

$$f_{\mu\nu}(x) = \partial_\mu X^A(x)\partial_\nu X^B(x) f_{AB}(X(x)) \quad (2.8)$$

so that the quantities X^A can be considered a set of 4 new dynamical scalar fields which will appear in the action. Considering some background solution for X_A, $X_0^A(x) \equiv \delta_\mu^A x^\mu$ we then introduce the "pion" fields π^A as

$$X^A(x) = X_0^A(x) + \pi^A(x) \quad (2.9)$$

We then further decompose the pion fields into scalar and vector Stueckelberg fields" $\hat{A}_\mu(x)$ and $\hat{\phi}(x)$:

$$\pi^A(x) = \delta_\mu^A(\hat{A}^\mu + \eta^{\mu\nu}\partial_\nu\hat{\phi}) \quad (2.10)$$

If we further expand around flat spacetime, writing $g_{\mu\nu} = \eta_{\mu\nu} + \hat{h}_{\mu\nu}$, we'll obtain an action in terms of three dynamical fields:

$$\hat{h}_{\mu\nu}(x), \;\; \hat{A}^{\mu}(x), \;\; \hat{\phi}(x) \tag{2.11}$$

Note that in the massless limit (to be defined below), all 5 degrees of freedom of the massive graviton will be captured in these fields: 2 degrees of freedom in the massless tensor $\hat{h}_{\mu\nu}$, 2 degrees of freedom in the massless vector $\hat{A}^{\mu}$, and 1 degree of freedom in a massless scalar $\hat{\phi}$.

A few more technical steps allow us to write the dRGT action in terms of $\hat{h}_{\mu\nu}$, $\hat{A}^{\mu}$, and $\hat{\phi}$. The quantity defined by $H_{\mu\nu} \equiv g_{\mu\nu} - f_{\mu\nu}$ can be written in terms of the Stueckelberg fields as

$$H_{\mu\nu} = \hat{h}_{\mu\nu} - \partial_{\mu}\hat{A}_{\nu} - \partial_{\nu}\hat{A}_{\mu} - 2\partial_{\mu}\partial_{\nu}\hat{\phi} - \partial_{\mu}\hat{A}_{\sigma}\partial_{\nu}\hat{A}^{\sigma} - \partial_{\mu}\partial_{\sigma}\hat{\phi}\partial_{\nu}\partial^{\sigma}\hat{\phi} - \partial_{\nu}\hat{A}^{\sigma}\partial_{\mu}\partial_{\sigma}\hat{\phi} - \partial_{\mu}\hat{A}^{\sigma}\partial_{\nu}\partial_{\sigma}\hat{\phi}.$$

Before proceeding, the fields should be canonically normalized as in [123]:

$$h_{\mu\nu} = M_P\hat{h}_{\mu\nu}, \qquad A_{\mu} = M_P m\hat{A}^{\mu}, \qquad \phi = M_P m^2\hat{\phi}. \tag{2.12}$$

Further, writing $\mathcal{K}$ as $\mathcal{K}^{\mu}_{\nu} = \delta^{\mu}_{\nu} - \sqrt{\delta^{\mu}_{\nu} - g^{\mu\alpha}H_{\alpha\nu}}$, one can then expand out the action (2.7) in terms of $h_{\mu\nu}$, $A^{\mu}(x)$, and $\phi(x)$, producing a slew of interaction terms.

In a generic non-linear massive gravity model, the leading interaction terms are of the form $\sim (\partial\phi)^3/(M_P m^4)$ [123]. The way that the dRGT theory is constructed is special in that these interactions $\sim (\partial\phi)^3/(M_P m^4)$ cancel, so that the new leading interaction term is suppressed by a higher scale:

$$\Lambda_3 = (m^2 M_P)^{1/3}. \tag{2.13}$$

This is accomplished by tuning the coefficients in the generic potential in a specific way. Thus, for the dRGT model, the *decoupling limit* is defined as:

$$M_P \to \infty, \qquad m \to 0, \qquad \Lambda_3 \sim \text{const}, \qquad \frac{T_{\mu\nu}}{M_P} \sim \text{const}. \tag{2.14}$$

In this decoupling limit, using the above definitions and a bit of algebra, the action obtained is:

$$S^{\text{decoup}}_{\text{dRGT}} = \int d^4x \left[-\frac{1}{2}h^{\mu\nu}\mathcal{E}^{\alpha\beta}_{\mu\nu}h_{\alpha\beta} + \frac{1}{2}h^{\mu\nu}X^{(1)}_{\mu\nu} - \frac{\alpha}{2\Lambda_3^3}h^{\mu\nu}X^{(2)}_{\mu\nu} + \frac{\beta}{2\Lambda_3^6}h^{\mu\nu}X^{(3)}_{\mu\nu} + T_{\mu\nu}h^{\mu\nu} \right] \tag{2.15}$$

where $\mathcal{E}_{\mu\nu} = \mathcal{E}^{\alpha\beta}_{\mu\nu}h_{\alpha\beta}$ is the linearization around $\eta_{\mu\nu}$ of the Einstein tensor $G_{\mu\nu}$ and we've simplified by combining coefficients: $\alpha \equiv -(1+\alpha_3)$, $\beta \equiv (\alpha_3 + \alpha_4)/3$. The tensors $X^{(n)}_{\mu\nu}$ are special conserved combinations of $\Phi \equiv \partial_{\mu}\partial_{\nu}\phi$ such that $\partial^{\mu}X^{(n)}_{\mu\nu} = 0$. They are defined as

$$X^{(1)}_{\mu\nu} = 2(\Phi_{\mu\nu} - [\Phi]\eta_{\mu\nu}) \tag{2.16}$$

$$X^{(2)}_{\mu\nu} = [\Phi]^2 - [\Phi^2])\eta_{\mu\nu} - 2([\Phi]\Phi_{\mu\nu} + \Phi^2_{\mu\nu}) \tag{2.17}$$

$$X^{(3)}_{\mu\nu} = ([\Phi]^3 - 3[\Phi][\Phi^2] + 2[\Phi^3])\eta_{\mu\nu} - 3([\Phi]^2 - [\Phi^2])\Phi_{\mu\nu} + 6[\Phi]\Phi^2_{\mu\nu} - 6\Phi^3_{\mu\nu}\,, \tag{2.18}$$

which can also be written using the Levi-Civita tensors,

$$\begin{aligned} X^{(1)}_{\mu\nu} &= \epsilon_\mu{}^{\alpha\rho\sigma}\epsilon_\nu{}^{\beta}{}_{\rho\sigma}\Phi_{\alpha\beta}, \\ X^{(2)}_{\mu\nu} &= \epsilon_\mu{}^{\alpha\rho\gamma}\epsilon_\nu{}^{\beta\sigma}{}_{\gamma}\Phi_{\alpha\beta}\Phi_{\rho\sigma}, \\ X^{(3)}_{\mu\nu} &= \epsilon_\mu{}^{\alpha\rho\gamma}\epsilon_\nu{}^{\beta\sigma\delta}\Phi_{\alpha\beta}\Phi_{\rho\sigma}\Phi_{\gamma\delta}. \end{aligned} \tag{2.19}$$

Square brackets are still used to denote the trace, $[\Phi] = \Phi_{\mu\nu}\eta^{\mu\nu}$. Notice that A^μ does not appear in the decoupling limit because it does not couple to a source, so it can be consistently set to zero. Due to the specific nature of this lagrangian, the coefficients α and β do not get renormalized by quantum loops, and thus their values are technically natural [124]. For generic values of these coefficients, the helicity-0 (ϕ) and helicity-2 ($h_{\mu\nu}$) modes still mix and cannot be completely decoupled. However, the mixing can be undone by choosing $\beta = 0$ [44], in which case there exists an invertible field redefinition,

$$h_{\mu\nu} \to h_{\mu\nu} + \eta_{\mu\nu}\phi + \frac{\alpha}{\Lambda_3^3}\partial_\mu\phi\partial_\nu\phi, \tag{2.20}$$

that decouples the scalar and tensor modes completely so that the Lagrangian can be written as $\mathcal{L} = \mathcal{L}_\phi + \mathcal{L}_h$ [44]. The remarkable property of the action (2.15) is that it represents the exact form of the action in the decoupling limit, meaning that all terms above quartic order vanish. Equation (2.15) represents the unique theory that any nonlinear, ghostless extension of massive gravity should reduce to in the decoupling limit [44].

It is important to emphasize the significance of the decoupling limit. First of all, the main purpose of the decoupling limit is to decouple the scalar degree of freedom from gravity as much as possible. In the DGP case that will be introduced in section 2.3, the resulting action (2.33) has π and $h_{\mu\nu}$ completely decoupled. In dRGT, the action (2.15) has ϕ and $h_{\mu\nu}$ mixed at all orders unless $\beta = 0$, but this is as decoupled as the action can get. In the decoupling limit, in both cases, the study of solutions is greatly simplified, thus, it is an obvious place to start.

An interesting feature of this limit is that it allows to decouple the scalar mode from gravity, while retaining the number of degrees of freedom of the theory, and keeping the full non-linear dynamics of the scalar field and its coupling to matter. Thus, within its range of validity, it describes important non-linear phenomena of the theory, such as the Vainshtein mechanism (see chapter 3). Another reason for working in this limit is due to the fact that we take the graviton mass to zero (2.14). Recall that a viable theory of massive gravity should agree with GR in the $m \to 0$ limit, at least within the solar system. Thus, in the decoupling limit, we can make a direct comparison of the massive gravity theory to GR.

It should also be noted that, in this limit, we are perturbing around flat space $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ with $|h_{\mu\nu}| \ll 1$ so the solutions are only valid in the weak field regime: for matter sources much

bigger than their Schwarzschild radius. Of course, the solutions found in the decoupling limit do not necessarily imply that solutions of the full theory exist with the same properties. There is the possibility that important properties of the full nonlinear system could be overlooked, and perhaps the decoupling limit cannot capture all of the interesting physics. As argued in [125], let us stress that the decoupling limit solutions give the local dynamics at scales within the present Hubble volume. Thus, solutions found in the decoupling limit should still appear as transients lasting for long cosmological times in the full theory [126].

2.2.3 Galileons

The above action (2.15) enjoys linear diffeomorphism invariance $\delta h_{\mu\nu} = \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$ (up to a total derivative). It is exactly invariant under the field space of *galileon* transformations $\delta\phi = c + b_\mu x^\mu$. In fact, the ϕ field is referred to as a *galileon* field [125]. With $\beta = 0$, and under the field redefinition (2.20), the decoupling limit lagrangian can be fully decoupled $\mathcal{L} = \mathcal{L}_\phi + \mathcal{L}_h$. $\mathcal{L}_h$ describes a typical GR massless graviton, and in the scalar sector, $\mathcal{L}_\phi$, we see the appearance of the galileon terms,

$$\mathcal{L}_\phi = -\frac{3}{2}\mathcal{L}_2(\phi) + \frac{3}{2}\frac{\alpha}{\Lambda_3^3}\mathcal{L}_3(\phi) - \frac{1}{2}\frac{\alpha^2}{\Lambda_3^6}\mathcal{L}_4(\phi) + \frac{\phi T}{M_4} + \frac{\alpha}{M_4\Lambda_3^3}\partial_\mu\phi\partial_\nu\phi T^{\mu\nu}\,, \tag{2.21}$$

where the galileon terms are given by

$$\mathcal{L}_2(\phi) = (\partial\phi)^2 \tag{2.22}$$

$$\mathcal{L}_3(\phi) = (\partial\phi)^2[\Phi] \tag{2.23}$$

$$\mathcal{L}_4(\phi) = (\partial\phi)^2([\Phi]^2 - [\Phi^2]) \tag{2.24}$$

$$\mathcal{L}_5(\phi) = (\partial\phi)^2([\Phi]^3 - 3[\Phi][\Phi^2] + 2[\Phi^3])\,, \tag{2.25}$$

where $(\partial\phi)^2 = \eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$ and $\Phi \equiv \partial_\mu\partial_\nu\phi$. These terms have two special properties:

1. Their equations of motion are always second order, despite the appearance of higher order derivatives in the lagrangian.

2. They are invariant up to a total derivative under the galileon symmetry $\phi(x) \to \phi(x) + c + b_\mu x^\mu$.

As shown in [125], these are the only polynomial terms in 4 dimensions with these properties. The fact that the equations of motion are always second order in derivatives is of key importance. If this were not the case, then by Ostrogradsky's theorem (see e.g. [127] for a clear exposition) the Hamiltonian would necessarily be unbounded from below, and stable solutions would not

exist. Essentially, this galileon symmetry ensures that the decoupling limit theory is ghost free. With some extra effort in [122], it was shown that the full theory beyond the decoupling limit, including all fields to all orders, carries just 5 degrees of freedom. Hence, remarkably, this Λ_3 theory is completely free of the Boulware-Deser ghost.

For the subspace of parameters defined by $\beta = 0$, the dRGT theory reduces to galileon-modified gravity with a quartic galileon. However, the decoupling limit of the dRGT theory does have some differences to simple galileon-modified gravity theories. First of all, when $\beta \neq 0$ there is an undiagonalizable interaction term, $\frac{\beta}{\Lambda_3^6} h^{\mu\nu} X^{(3)}_{\mu\nu}$ and a non traditional coupling $\nabla^\mu \phi \nabla^\nu \phi T_{\mu\nu}$, the effects of which will be studied in section 3.7.

Galileons were first discovered in the DGP model, introduced in the next section, in which the cubic galileon $\mathcal{L}_3$ describes the leading interactions of the scalar mode [128, 129]. The properties of galileons have since been abstracted away from DGP. They have taken on a life of their own as they've been shown to have several interesting properties, such as a non-renormalization theorem [130], superluminal propagation [131, 132], galileon dualities [133, 134], and extended symmetries [135, 136].

2.3 The DGP Model

Massive gravity has also been studied in the context of extra-dimensions in the Dvali-Gabadadze-Porrati model (DGP) [137]. The connection between the DGP model and massive gravity can be related to the fact that in the DGP model, gravity is mediated by a continuum of massive Kaluza-Klein gravitons due to the higher dimensional nature of the bulk spacetime. The DGP model is also a subset of the galileon theories. Historically, the DGP model played the most important role for the recent developments of massive gravity. The starting point for the analyses in chapter 3 is in fact the decoupling limit of the DGP model in flat space.

The DGP model [137] has surely been one of the most historically important models for developing massive gravity theories. It is a 5-dimensional brane-world model, and it describes our 4-dimensional universe as a surface embedded into a 5D bulk space-time. One can integrate out the bulk degrees of freedom and find an "effective" action for the 4D fields. In [128] it was shown that from the 4D point of view, besides the ordinary graviton, an extra scalar degree of freedom plays a crucial role. This extra scalar field will be denoted as π. It represents a brane-bending mode, meaning that it contributes to the extrinsic curvature of the boundary as $K_{\mu\nu} \propto \partial_\mu \partial_\nu \pi$. All interesting physics, good or bad, of the DGP model can be traced to the dynamics of the π scalar.[1] So for our purposes, we'll forget about the 5D geometric setup of

[1] In the decoupling limit of dRGT massive gravity, we denote the scalar field as ϕ. In the decoupling limit

the model and simply study the scalar 4D theory. This is justified since all physics at distances less than H^{-1} is correctly reproduced by the π scalar coupled to 4D gravity, without any reference to the theory's 5D origin.

This model is interesting for several reasons. It has been shown that it provides an alternative explanation for the observed recent accelerated expansion of the universe [138] through solutions that are self-accelerating. Another key interest in the DGP model lies in its successful implementation of the Vainshtein mechanism [139] which is able to decouple the additional modes from gravitational dynamics at shorter distances. The Vainshtein mechanism, which will be thoroughly introduced in chapter 3, screens the contribution of π below the experimentally acceptable level [140] without introducing extra ghost degrees of freedom. We'll also see that the π field in the DGP model is in fact a simple galileon field with leading interaction $\mathcal{L}_3$, and so enjoys all of the properties shared by galileons (most importantly that the equations of motion are second order in the derivatives).

The DGP model describes gravity in a 5D spacetime $\mathcal{M}$ with 4D boundary $\partial\mathcal{M}$ and is postulated to have the following action:

$$S_{\rm DGP} = 2M_5^3 \int_{\mathcal{M}} d^5x\sqrt{-G}R^{(5)}(G) + 2M_4^2 \int_{\partial\mathcal{M}} d^4x\sqrt{-g}R^{(4)}(g) - 4M_5^3 \int_{\partial\mathcal{M}} d^4x\sqrt{-g}K(g),$$

where G_{MN} is the 5D metric, $g_{\mu\nu}$ is the 4D metric induced on $\partial\mathcal{M}$, and $K(g)$ is the Gibbons-Hawking term on $\partial\mathcal{M}$ which is added so that the 5D Einstein equations are obtained upon variation of the bulk action [141]. There is a special length scale

$$L_{\rm DGP} = \frac{1}{m} \equiv \frac{M_4^2}{M_5^3}, \tag{2.26}$$

where M_4, M_5 are the 4 and 5 dimensional Planck masses. We often denote $M_4 \equiv M_P$ when it is understood that we are working in 4D. Below $L_{\rm DGP}$ the theory looks 4 dimensional, while above it, the theory enters the 5D regime. To make the analogy to massive gravity clear, m is considered as the graviton mass.

This will be an overview of the results obtained in [128], where the 4D boundary effective action is obtained by integrating out the bulk degrees of freedom. Using spacetime coordinates $x^M = (x^\mu, y)$ with the boundary at $y = 0$, the bulk part of the action can be written in the ADM language [120] as,

$$S_{\rm bulk} = 2M_5^3 \int d^4x \int_0^\infty dy\sqrt{-g}N[R(g) - K_{\mu\nu}K^{\mu\nu} + K^2], \tag{2.27}$$

where $N \equiv 1/\sqrt{G^{yy}}$ is the lapse, $N_\mu \equiv G_{y\mu}$ is the shift, and the extrinsic curvature tensor is $K_{\mu\nu} = \frac{1}{2N}(\partial_y g_{\mu\nu} - \nabla_\mu N_\nu - \nabla_\nu N_\mu)$. The next step is to expand this around a flat background,

of DGP, the scalar field is traditionally denoted as π. They play the same role as the helicity-0 mode of the graviton, and essentially a galileon modification to GR in the decoupling limit.

$G_{MN} = \eta_{MN} + h_{MN}$, and integrate out the bulk to obtain an effective action for the 4D fields on the boundary. The result at quadratic level is

$$S_{\rm bdy} = M_4^2 \int d^4x \left[\frac{1}{2} h'^{\mu\nu} h'_{\mu\nu} - \frac{1}{4} h' \Box h' - m N'^{\mu} \Delta N'_{\mu} + 3m^2 \hat{\pi} \Box \hat{\pi} \right], \tag{2.28}$$

where $\Delta = \sqrt{-\Box} = \sqrt{-\eta^{\mu\nu}\partial_\mu \partial_\nu}$, and the following definitions have been made to diagonalize the kinetic terms:

$$h_{yy} = -2\Delta\hat{\pi}, \qquad N'_\mu = N_\mu - \partial_\mu \hat{\pi}, \qquad h'_{\mu\nu} = h_{\mu\nu} - m\hat{\pi}\eta_{\mu\nu}. \tag{2.29}$$

Going past the quadratic, level, taking into account higher powers of h_{MN}, one finds that the leading boundary interaction term is cubic in $\hat{\pi}$ and has the form $-M_5^3 \int d^4x (\partial\hat{\pi})^2 \Box\hat{\pi}$. Now compare the coefficient of this interaction term with the kinetic term in $S_{\rm bdy}$, and one immediately sees that the theory becomes strongly interacting at the energy scale

$$\Lambda_{DGP} = (m^2 M_4)^{1/3} = \frac{M_5^2}{M_4}, \tag{2.30}$$

which is equivalent to the Λ_3 scale in the dRGT theory, and corresponds to a length scale of about 1000 km. From now on we'll just denote this scale as Λ_3. It is the lowest scale of all strong interaction scales associated to further interaction terms. It is verified in [129] that all other interaction terms are subdominant as long as a flat-space approximation is taken: $|h'_{\mu\nu}| \ll 1$.

Let's consider this flat-space approximation in more detail. The region where space is nearly flat outside of a compact source is well outside its Schwarzschild radius R_S, so we can pretend we are in this region by sending $R_S \to 0$ which would decouple 4D gravity from the source. Care needs to be taken so that the resulting configuration is not simply trivial $h'_{\mu\nu} = 0$ and the self-couplings of the π field are preserved along with its coupling to the source T. If there is the usual coupling of $h_{\mu\nu}$ and the energy-momentum tensor $T_{\mu\nu}$ given by $\frac{h_{\mu\nu}T^{\mu\nu}}{2}$, then one can see from the definition of $h'_{\mu\nu}$ in (2.29) that a coupling $\frac{1}{2M_4}\pi T$ is induced, where we have canonically normalized the scalar field:

$$\pi = M_4 m \hat{\pi} \tag{2.31}$$

Since we want to keep the π self-couplings (with coefficient $\sim \Lambda_3$) and the coupling of π to T (with coefficient $\sim \frac{T}{M_4}$) in the flat-space/Minkowski approximation, the decoupling limit in the DGP theory is given by:

$$M_4, M_5, T_{\mu\nu} \to \infty, \qquad \Lambda_3 \sim \text{const}, \qquad \frac{T_{\mu\nu}}{M_4} \sim \text{const}. \tag{2.32}$$

This limit is analogous to the decoupling limit in the dRGT theory (2.14). In this limit, the 4D graviton is decoupled while the full lagrangian for π (kinetic, cubic and source terms) is

kept fixed. The full action for the π field in flat space in this decoupling limit is given by:

$$S_\pi = \int d^4x \left[-3(\partial\pi)^2 - \frac{1}{\Lambda_3^3}(\partial\pi)^2\Box\pi + \frac{1}{2M_4}\pi T \right]. \tag{2.33}$$

One can recognize appearance of the quadratic (2.22) and cubic (2.23) galileon terms above. Explicitly,

$$\mathcal{L}_\phi = -3\mathcal{L}_2(\pi) - \frac{1}{\Lambda_3^3}\mathcal{L}_3(\pi) + \frac{\pi T}{2M_4}. \tag{2.34}$$

The galileon symmetry ensures that the equations of motion for π are second order in derivatives, and therefore no Boulware-Deser ghost propagates at the background level. However, ghost instabilities can appear at the nonlinear level if the kinetic term for perturbations on top of a background acquire the wrong sign with reference to other fluctuating degrees of freedom. This was the main obstruction to obtaining a viable explanation for the observed cosmological accelerated expansion in the context of DGP [128, 129, 142, 143, 144].

2.4 Bigravity

Part of the motivation for going beyond massive gravity is discovered when one attempts to find cosmological solutions within the dRGT theory (see [145] for a review). Under a standard FRW ansatz for the metric $g_{\mu\nu}$, the equations of motion that result from the action (2.6) are immediately problematic, implying that the only solution for the scale factor is a constant [146]. Therefore, no flat homogeneous and isotropic cosmological solutions exist within dRGT massive gravity. The only FRW solutions that do exist are open. There are two such branches: a normal branch which is just Minkowski in a different slicing, and a self-accelerating branch, which describes an accelerating cosmology (without a cosmological constant) with scale $H \sim m$ set by the graviton mass [146, 126, 147, 148]. However, when perturbations around the self-accelerating branch are analyzed, it was revealed that the scalar and vector modes have vanishing kinetic terms [149, 150], implying that the solutions are infinitely strongly coupled, and beyond the analysis of perturbative techniques. This occurs for both flat and FRW reference metrics.

If one hopes to examine cosmology in the context of dRGT theories, the assumptions of homogeneity and isotropy must be abandoned. Although some anisotropic solutions with well behaved perturbations have been found [151, 152], analyzing the cosmological implications of these models is challenging since most of our standard cosmological tools rely on homogeneity and isotropy. Therefore, most work in massive gravity cosmology has stuck with the attempt to find exact FRW solutions, requiring instead a generalization of the original dRGT theory. A particularly popular and natural generalization is found in bigravity, which will now be demonstrated.

As noted in equation (2.3), in trying to non-linearly extend the Fierz-Pauli theory, the necessary mass term in the action requires the introduction of a fixed non-dynamical tensor field, $f_{\mu\nu}$, in addition to the metric $g_{\mu\nu}$ describing spacetime [117]. The dependence of the theory on an arbitrary background metric is not theoretically desirable. Promoting the fixed tensor $f_{\mu\nu}$ to a dynamical field, a ghost-free theory of *massive bigravity* [153, 154] has recently been introduced. With two dynamical metrics, the choice of matter coupling is a non-trivial issue [155, 156]. The only consistent choice that does not introduce a ghost is to have couplings of matter to one metric only. In this scenario, one metric $g_{\mu\nu}$ describes our spacetime, and another one, $f_{\mu\nu}$, is part of a "dark" gravitational sector.

Starting from the dRGT action in equation (2.6), we will consider the natural extension to massive bigravity proposed by Hassan and Rosen [157]:

$$\begin{aligned} S_{\rm bi} = &- \frac{M_g^2}{2}\int d^4x\sqrt{-g}R(g) - \frac{M_f^2}{2}\int d^4x\sqrt{-\tilde{f}}R(\tilde{f}) \\ &+ m^2M_g^2\int d^4x\sqrt{-g}\sum_{n=0}^{4}\tilde{\beta}_n e_n(\sqrt{g^{-1}\tilde{f}}) + S_{\rm matter}\,, \end{aligned} \tag{2.35}$$

where the two dynamical metrics are $g_{\mu\nu}$ and $\tilde{f}_{\mu\nu}$ and their respective Planck masses are M_g and M_f, while m is the mass scale associated with the graviton mass matrix. The interaction term between g and $\tilde{f}$ contains a linear combination of the symmetric polynomials $e_n(X)$ given in equation (2.5). The dimensionless coefficients $\tilde{\beta}_n$ are free parameters of the theory. Further, we will consider a singly-coupled theory in which $S_{\rm matter}$ contains only couplings of matter to $g_{\mu\nu}$. In this case, $g_{\mu\nu}$ is considered the standard physical metric while $\tilde{f}_{\mu\nu}$ is a new dynamical tensor field. This singly-coupled theory was shown to be free of the Boulware-Deser ghost [155]. The resulting quantum corrections (at one loop) are nothing other than the standard cosmological constant which does not detune the special structure of the graviton potential and is thus harmless [155]. By rescaling the dark metric and the free parameters as follows,

$$f_{\mu\nu} = (M_f/M_g)^2\tilde{f}_{\mu\nu} \qquad \beta_n^* = (M_g/M_f)^n\tilde{\beta}_n\,, \tag{2.36}$$

we can rewrite the action so that the redundant scale M_f is absent:

$$\begin{aligned} S_{\rm bi} = &M_g^2\left[-\frac{1}{2}\int d^4x\sqrt{-g}R(g) - \frac{1}{2}\int d^4x\sqrt{-f}R(f)\right. \\ &\left. + m^2\int d^4x\sqrt{-g}\sum_{n=0}^{4}\beta_n^* e_n(\sqrt{g^{-1}f}) + \frac{S_{\rm matter}}{M_g^2}\right]. \end{aligned} \tag{2.37}$$

It has been shown that the cosmological expansion and spherically-symmetric solutions to this theory give viable alternatives to general relativity [158, 159, 160, 161, 162], at the background level. Let us briefly examine the background equations of motion and solutions.

2.4.1 Bigravity cosmology

We start by making an FRW ansatz for the metrics $g_{\mu\nu}$ and $f_{\mu\nu}$:

$$ds_g^2 = a^2(\tau)(-d\tau^2 + dx_i dx^i), \tag{2.38}$$

$$ds_f^2 = b^2(\tau)[-c^2(\tau)d\tau^2 + dx_i dx^i], \tag{2.39}$$

where τ represents conformal time, a and b are the scale factors corresponding to the g and f metric respectively, and c is the lapse function for the f metric. Note that the assumption that $f_{\mu\nu}$ is FRW is not the most general choice for the metric, and it could have more complicated dynamics.

The Bianchi identities imply the following relation:

$$c = \frac{\hat{\mathcal{H}}_f}{\hat{\mathcal{H}}} = \frac{\dot{b}a}{\dot{a}b} = 1 + \frac{\dot{r}}{r\hat{\mathcal{H}}}. \tag{2.40}$$

where $\hat{\mathcal{H}} = \dot{a}/a$ is the conformal Hubble function for the g metric, $\hat{\mathcal{H}}_f = \dot{b}/b$ is the conformal Hubble function for the f metric, and $r = b/a$ the ratio of scale factors. For simplicity we perform the following rescaling

$$\beta_n = \frac{m^2}{H_0^2}\beta_n^*, \qquad \bar{\rho} = \frac{\rho}{M_g^2 H_0^2}, \qquad \mathcal{H} = \frac{\hat{\mathcal{H}}}{H_0} \tag{2.41}$$

where $\bar{\rho} = \bar{\rho}_m + \bar{\rho}_r$ is the dimensionless energy density of all matter and radiation components, and we measure all times and lengths in terms of H_0. With these definitions in hand, variation of the action with respect to $g_{\mu\nu}$, and inserting (2.38) and (2.39), gives the Friedmann equations

$$3\mathcal{H}^2 = a^2\left(\bar{\rho} + \bar{\rho}_{\rm mg}\right), \tag{2.42}$$

$$2\dot{\mathcal{H}} + \mathcal{H}^2 = a^2\bar{\rho} + \frac{a^3}{3}\frac{d\bar{\rho}}{da} + a^2\left(\beta_0 + \beta_1 r(2+c) + \beta_2 r^2(1+2c) + \beta_3 r^3 c)\right), \tag{2.43}$$

$$\text{where} \qquad \bar{\rho}_{\rm mg} \equiv \left(\beta_0 + 3\beta_1 r + 3\beta_2 r^2 + \beta_3 r^3\right), \tag{2.44}$$

where $\bar{\rho}_{\rm mg}$ is an effective massive-gravity energy density. We also have the background equations for the f metric:

$$3\mathcal{H}^2 = \frac{a^2}{r}\left(\beta_1 + 3\beta_2 r + 3\beta_3 r^2 + \beta_4 r^3\right), \tag{2.45}$$

$$2\dot{\mathcal{H}} + \mathcal{H}^2 c = \frac{a^2}{r}\left(\beta_1 + \beta_2 r(2+c) + \beta_3 r^2(1+2c) + \beta_4 r^3 c\right), \tag{2.46}$$

The energy densities follow the usual conservation laws

$$\dot{\bar{\rho}}_m + 3\mathcal{H}\bar{\rho}_m = 0 \quad \text{and} \quad \dot{\bar{\rho}}_r + 4\mathcal{H}\bar{\rho}_r = 0\ , \tag{2.47}$$

giving rise to their solutions in terms of a and the present day density parameters $\Omega_{i,0} = \rho_{i,0}/(3H_0^2M_g^2) = \bar{\rho}_{i,0}/3$,

$$\bar{\rho}_m = \frac{\Omega_{m,0}}{3a^3} \quad \text{and} \quad \bar{\rho}_r = \frac{\Omega_{r,0}}{3a^4}. \tag{2.48}$$

The above equations can be combined to form convenient equations for the dynamic variables r and a:

$$0 = \bar{\rho}_m + \bar{\rho}_r - \frac{1}{r}\left(\beta_1 + 3\beta_2 r + 3\beta_3 r^2 + \beta_4 r^3\right) + \beta_0 + 3\beta_1 r + 3\beta_2 r^2 + \beta_3 r^3, \quad (2.49)$$

$$\frac{1}{r}\frac{dr}{da} = -\frac{3}{a}\frac{\beta_3 r^4 + (3\beta_2 - \beta_4)r^3 + 3(\beta_1 - \beta_3)r^2 + (\beta_0 - 3\beta_2)r - \beta_1 - \Omega_{r,0}a^{-4}r}{3\beta_3 r^4 + 2(3\beta_2 - \beta_4)r^3 + 3(\beta_1 - \beta_3)r^2 + \beta_1}. \quad (2.50)$$

The procedure now is as follows:

1. Fix the free parameters: $\Omega_{m,0},\ \Omega_{r,0},\ \beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4$.
2. Use (2.49) evaluated today to find r_0, the present value of r.
3. Starting with initial condition r_0, evolve (2.50) to find $r(a)$ for all time.
4. Using $r(a)$, solve for the Hubble parameter $\mathcal{H}(a)$ using (2.45).
5. It is now possible to find $a(\tau)$ using $\mathcal{H} = \dot{a}/a$, and thus can also find $r(\tau)$ and $\mathcal{H}(\tau)$ and $c(\tau)$

In this work we will focus on two distinct types of background solutions, following the notation in [163]. A plot of $r(a)$ in each branch is shown in figure 2.1.

1. **The Expanding Branch**: In this branch of solution, both metrics g and f expand with time. This is also known as the finite branch since the ratio $r = b/a$ evolves from zero at early times to a finite value. Within this branch, there is a proposed minimal model [164, 159] in which only $\beta_1 \neq 0$. At the background level, it was shown that this theory can be compatible with expansion histories, but remains distinct from GR with testably different observables [164, 159]. The issue with this branch is an exponential perturbative instability in the scalar sector, previously noted in the literature [165, 166, 167, 163]. We therefore focus on this branch as an example of how phenomenologically different the results for tensors can be under different assumptions about the background solution. We will work within the minimal model, fixing $\beta_1 = 1.38$, which was found from a best fit analysis at the background level in [159]. In this case, an analytic solution exists for $r(a)$ and is given by

$$r(a) = \frac{-3a\Omega_{m,0} - 3\Omega_{r,0} + \sqrt{3(4a^8\beta_1^2 + 3a{\Omega_{m,0}}^2 + 6a\Omega_{m,0}\Omega_{r,0} + 3{\Omega_{r,0}}^2)}}{6a^4\beta_1}. \quad (2.51)$$

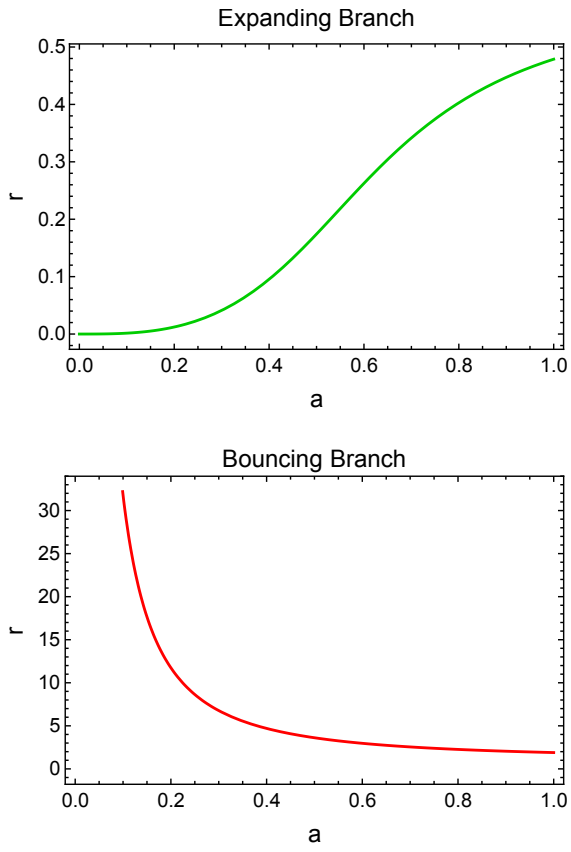

Figure 2.1: The evolution of the ratio of scale factors $r = b/a$ for the both background branches described in the text.

2. **The Bouncing Branch**: This is a more exotic option in which the physical metric g expands in time while the dark sector metric f experiences a bounce. At the bounce point, $f_{00} = 0$, thus, $f^{-1}_{\mu\nu}$ diverges, but there is no divergence in the physical sector [163, 168, 169]. In addition, well-defined and stable solutions for the background and linear perturbations exist through this point [163] indicating that this divergence is likely of a mathematical rather than physical nature. The perturbative instability of the expanding branch is absent in this sector. This branch is also called the infinite branch since r evolves from infinity at early times to a finite value at late times. Here we are required to set $\beta_0 = \beta_2 = \beta_3 = 0$, and for the remaining parameters, we use the best-fit values $\beta_1 = 0.48$ and $\beta_4 = 0.94$ found by fitting growth histories and type Ia supernovae [167].

Chapter 3

Vainshtein screening in massive gravity

This chapter explores the dynamics of the Vainshtein screening mechanism in the decoupling limit of the DGP braneworld scenario and dRGT massive gravity. This mechanism must work effectively to shield the extra degrees of freedom, providing agreement with GR on small scales. Otherwise, traditional tests of gravity, such as solar system tests, would be enough to rule out the theory, similar to the vDVZ discontinuity that plagued Fierz-Pauli massive gravity. Our work here builds upon the known static screening solutions in the decoupling limit of massive gravity, and addresses their dynamical evolution and stability. We show that there is a vast set of initial conditions whose evolution is well defined and which are driven to the static screening solutions. Screening solutions are stable and behave coherently under small fluctuations: they oscillate and eventually settle to an equilibrium configuration, the timescale for the oscillations and damping being dictated by the Vainshtein radius of the screening solutions. At very late times, a power-law decay ensues, in agreement with known analytical results. However, we also conjecture that physically interesting processes such as gravitational collapse of compact stars may not possess a well-posed initial value problem. Finally, we construct solutions with nontrivial multipolar structure describing the screening field of deformed, asymmetric bodies and show that higher multipoles are screened more efficiently than the monopole component.

3.1 Introduction

The strongest motivation for studying massive gravity is the possibility of an alternative explanation to cosmic acceleration. In the standard cosmological model, the accelerated expansion is assumed to be due to a cosmological constant. However, the extreme fine tuning of the cosmological constant has led physicists to pursue alternatives. If the mass of the graviton is on the order of the Hubble constant, $m \sim H_0 \sim 10^{-33}$ eV, massive gravity admits *self-accelerating* solutions which produce a universe undergoing accelerated expansion without explicitly using a cosmological constant. It is through the new dynamics of the theory, rather than through a cosmological constant, that we are able to achieve cosmic acceleration in massive gravity. In this case, the tuning of the mass m is technically natural, meaning that it is stable to quantum corrections. This coincides with the fact that the theory gains an additional symmetry in the $m \to 0$ limit (the gauge symmetry of GR). This is therefore an improvement over the fine tuning of the cosmological constant which is not technically natural, and is unstable to quantum corrections.

To provide an explanation for accelerated expansion, massive gravity should be an infrared modification to GR, but the theories should agree on solar system scales where we know that GR is successful. It is required that the extra degrees of freedom in massive gravity shield themselves on these short distance scales. This is achieved in massive gravity via the *Vainshtein screening mechanism* [139], a mechanism that relies on non-linear derivative couplings in the theory to screen the extra degrees of freedom at short distances, and thus plays a crucial role for phenomenological applications of massive gravity.

The Vainshtein mechanism was first studied for its role in establishing the consistency between the massless limit of the Fierz-Pauli massive graviton [42] and GR [115, 170]. This mechanism is also exhibited in theories with galileon symmetry [125] as first discovered in the Dvali-Gabadadze-Porrati (DGP) braneworld model [137]. Within the context of DGP and dRGT massive gravity, extensive work has been done identifying static spherically symmetric solutions which manifest the Vainshtein mechanism [129, 171, 172, 173, 174, 175, 176, 177, 178]. This is an important verification on the compatibility of these models with solar system tests, and provides insight into how the evolution of the universe is affected on large scales.

Due to the complex structure of the non-linear equations of motion in these theories, there has been little exploration of fully dynamical phenomena. There has also been no study of solutions without spherical symmetry. Some notable exceptions are the study of scalar gravitational radiation emitted by a binary pulsar system [179, 180], computations of the Green's function [181], plane wave solutions [182], studies of the characteristic matrix of the

full dRGT theory [183, 184, 185, 186], and some results on stationary solutions [187]. In this chapter [1], we fill some of the existing gaps in the knowledge of fully dynamical and aspherical solutions in theories displaying the Vainshtein mechanism. We study two theories in the decoupling limit: DGP and dRGT gravity. In the decoupling limit, a non-gravitating scalar degree of freedom is introduced that couples to the matter sector. We consider non-relativistic sources and neglect the internal dynamics of the matter sector. Under these approximations, we consistently solve for the full dynamical evolution of the scalar on a fixed background Minkowski space with fixed sources. The two theories differ primarily in their derivative self-couplings and couplings to matter, but qualitatively the dynamical behaviour is similar. Our results represent a first step towards establishing the nonlinear dynamical stability of infrared modifications of gravity.

Following an overview of the Vainshtein mechanism in section 3.2, we begin by characterizing the linear and non-linear stability of the screening solutions in the DGP model in section 3.3. Using numerical simulations, we find that in the fully non-linear regime a large space of initial conditions evolve to the screening solution around a fixed time-independent source at sufficiently late times. We study the linear response properties of the screening profiles, showing that the screening solutions behave much as a black hole when perturbed: there is a prompt response, followed by a universal ring-down and power-law decay of the wave form. We also examine the screening properties of static non-spherically symmetric scalar profiles and sources.

However, we also find that for initial data sufficiently far from the screening solution, the evolution becomes ill-posed: regions form where the constant time surfaces fail to be Cauchy surfaces of the scalar equation of motion. Following Ref. [188], we term this Cauchy breakdown. We study the situations in which Cauchy breakdown occurs in section 3.4, and show that no foliation of flat spacetime exists in which the evolution can be continued.

In addition to static sources, in section 3.5, we consider two models of a dynamical source corresponding to spherically symmetric collapse and explosion. However, in both cases, for sufficiently dense sources or sufficiently short timescales of collapse, we find that Cauchy breakdown occurs. These examples suggest that a fully dynamical study of astrophysical phenomena for infrared modifications of gravity will require a prescription for evolving past Cauchy breakdown, or an ultraviolet completion of the theory. This work motivates a more complete treatment which will require stepping away from the decoupling limit, and considering realistic dynamical sources.

In section 3.6 we generalize our analysis to asymmetric screening solutions. In section 3.7, we repeat the full dynamical stability analysis for the dRGT model. A summary of results is

presented in section 3.8.

3.2 Vainshtein mechanism

The Vainshtein mechanism relies on derivative self-interactions to screen long-range fifth forces mediated by a scalar degree of freedom. There are a variety of theories that manifest the Vainshtein mechanism including the DGP braneworld scenario, dRGT massive gravity, massive bi-metric gravity, and scalar field theories with Galileon invariance. In each case, the action contains higher-derivative interaction terms for the fluctuating degrees of freedom, but the equations of motion remain second order.

In this work, we restrict our focus to DGP and dRGT massive gravity. The DGP model [137] physically describes our universe as a 3-brane embedded in a 5D bulk, introducing a brane-bending mode that from the 4D point of view corresponds to an additional scalar degree of freedom. dRGT massive gravity [43, 44] is the non-linear generalization of the Fierz-Pauli massive graviton. The theory propagates 5 degrees of freedom, two tensor, two vector, and one scalar; we focus on the physics of the scalar sector. Attempts have been made to explain the observed accelerated expansion of the universe in the context of both models (or close cousins), and one might expect that many of the features present in these models will be shared by other infrared modifications of gravity relying on similar non-linear derivative interactions.

Both DGP and a sector of dRGT massive gravity are special cases of a general effective theory with action:

$$S = \int d^4x \; \mathcal{L}_\pi \,, \tag{3.1}$$

where

$$\mathcal{L}_\pi = c_2\mathcal{L}_2 + \frac{c_3}{\Lambda_3^3}\mathcal{L}_3 + \frac{c_4}{\Lambda_3^6}\mathcal{L}_4 + \frac{c_5}{\Lambda_3^9}\mathcal{L}_5 + \frac{\xi\pi T}{M_4} + \frac{\alpha}{M_4\Lambda_3^3}\partial_\mu\pi\partial_\nu\pi T^{\mu\nu} \,, \tag{3.2}$$

$\Pi_{\mu\nu} \equiv \nabla_\mu\partial_\nu\pi$; c_2, c_3, c_4, c_5, α, ξ are dimensionless parameters; Λ_3 is the strong coupling scale of the theory; $\mathcal{L}_n$ are the galileon lagrangians given in (2.22)-(2.25); $T^{\mu\nu}$ is the energy momentum tensor of the source, and $T = \eta_{\mu\nu}T^{\mu\nu}$ is the trace.

Throughout this chapter, we restrict our focus to the decoupling limit, in which there is no direct coupling between the scalar and gravitational sectors (note the lack of minimal coupling to gravity in eq. 3.1). See section 2.2.2 to review the key features of the decoupling limit. The Lagrangian eq. 3.2 is the most general effective theory displaying the Vainshtein mechanism with no couplings between the scalar and gravitational sectors that can be obtained from the Horndeski action [189] in the decoupling limit. A restricted sector (with $\beta = 0$) of the decoupling limit of dRGT massive gravity [44] is recovered by setting $c_2 = -3/2$, $c_3 = 3\alpha/2$,

$c_4 = -\alpha^2/2$, $c_5 = 0$, $\xi = 1$ in the Lagrangian (3.2), as shown in eq. (2.34). The decoupling limit of the DGP model [129] can be obtained by setting $c_2 = -3$, $c_3 = -1$, $\xi = 1/2$, $c_4 = c_5 = \alpha = 0$, as shown in eq. (2.21).

An important length scale in these theories is the Vainshtein radius R_V, below which the non-linearities in the equation of motion become important. These non-linearities are crucial for the Vainshtein mechanism. Continuity with GR is recovered on length scales less than R_V by screening the scalar degree of freedom. This is crucial for phenomenological applications of modified gravity models. The Vainshtein radius is given by

$$R_V \equiv \frac{1}{\Lambda_3}\left(\frac{M(r\to\infty)}{M_4}\right)^{1/3}\,, \tag{3.3}$$

and retains its form in the decoupling limit as can be seen from (2.32). Here, M is the mass of the source defined by

$$M(r) \equiv -4\pi\int_0^r dr' r'^2 T(r')\,. \tag{3.4}$$

It will be convenient to measure all dimensionful quantities in terms of Λ_3 and M_4. Explicitly, we will use the following rescaled quantities

$$x^\mu\Lambda_3 \to x^\mu, \quad \frac{\pi}{\Lambda_3}\to\pi, \quad \frac{T}{M_4\Lambda_3^3}\to T, \tag{3.5}$$

in which case,

$$R_V = (M(r\to\infty))^{1/3}\,. \tag{3.6}$$

It can be checked from the equation of motion (3.8) that this is equivalent to working with the dimensionless density $\rho = \rho_d/(M_4\Lambda_3^3)$ where ρ_d is the dimensionful density. For a sun-like star, $\rho_d \sim 1.4\times10^3\text{kg/m}^3 \sim 5.9\times10^{-18}\text{GeV}^4$ and the strong coupling scale is $\Lambda_3 \sim M_5^2/M_4 \sim (1000\text{km})^{-1} \sim 1.8\times10^{-21}\text{GeV}$. Using $M_4 = 2.4\times10^{18}\text{GeV}$ we get the dimensionless density, to be used in our adopted units, of $\rho \sim 4.2\times10^{26}$.

For much of this work, we consider a simple spherically symmetric non-relativistic source with central density ρ and radius R_0 described by

$$T_{\mu\nu} = \text{diag}\left(\rho\exp\left(-\frac{r^2}{R_0^2}\right), 0, 0, 0\right)\,, \tag{3.7}$$

Note that the source is exponentially close to zero for $r \gtrsim R_0$. The Vainshtein radius for this source is given by $R_V = (M(r\to\infty))^{1/3} = \sqrt{\pi}\rho^{1/3}R_0$. For simplicity we do not treat the internal dynamics of the source, nor are we allowing it to evolve in the presence of the scalar field. This is a convenience, and not a physical choice. Nevertheless, we expect our results to be very weakly dependent on the precise radial distribution of $T(r)$ and future work will incorporate the full dynamics of a realistic source.

One can also consider a source with non-zero pressure, which would have important effects on the solution and its stability. For an increasingly relativistic source, the trace of the energy momentum tensor shrinks, weakening the coupling between the source and the scalar (this has been shown to have important implications for the Chameleon screening mechanism [190]). As another example, in dRGT massive gravity the disformal coupling $\nabla_\mu\phi\nabla_\nu\phi T^{\mu\nu}$ can give rise to instabilities whenever $p \sim \rho$ [191].

3.3 Screening in the Decoupling Limit of DGP

3.3.1 DGP equations of motion

We will first consider the DGP model, which is the simplest setting in which the Vainshtein mechanism arises. The full action for the scalar field in the DGP decoupling limit has already been expressed in equations (2.33). Varying this action gives rise to the following equation of motion (in our rescaled variables (3.5)),

$$6\Box\pi + 2\left(\Box\pi\right)^2 - 2\left(\nabla^\mu\nabla^\nu\pi\right)\left(\nabla_\mu\nabla_\nu\pi\right) + \frac{T}{2} = 0\,, \tag{3.8}$$

where the metric is included to allow for easy conversion between alternative foliations of Minkowski space. Under the assumption of spherical symmetry, and applying the rescaling given in (3.5), the equation of motion (3.8) for $\pi(r,t)$ can be written as

$$\ddot{\pi}\left(6 + 8\frac{\pi'}{r} + 4\pi''\right) = \frac{T}{2} + 12\frac{\pi'}{r} + 6\pi'' + 4(\dot{\pi}')^2 + 8\frac{\pi'\pi''}{r} + 4\frac{(\pi')^2}{r^2}\,, \tag{3.9}$$

where $\dot{\pi} = \partial\pi/\partial t$ and $\pi' = \partial\pi/\partial r$. General solutions to this non-linear partial differential equation are not available, and a numerical analysis is required (see Section 3.3.2 below). Nevertheless, static solutions – which establish the viability of the Vainshtein mechanism – can be obtained analytically. Dropping the time-dependence in eq. (3.9) allows us to integrate it to obtain a simple algebraic relation for $\Pi \equiv \pi'$

$$4\pi r^2\left(6\Pi(r) + 4\frac{\Pi(r)^2}{r}\right) = \frac{M(r)}{2}. \tag{3.10}$$

This quadratic equation has two solutions given by

$$\Pi_\pm(r) = -\frac{3r}{4} \pm \sqrt{\frac{M(r)}{32\pi} + \left(\frac{3r}{4}\right)^2} \tag{3.11}$$

which can be integrated once more to obtain the field $\pi_\pm(r)$. The scalar fifth force is proportional to $\Pi_\pm(r)\hat{r}$, and comparing with the Newtonian gravitational force on different

scales one obtains:

$$\left|\frac{F_{\Pi_+}}{F_g}\right| \sim \frac{\Pi_+/M_4}{M/(M_4^2 r^2)} \sim \begin{cases} \left(\frac{r}{R_V}\right)^{3/2} & \text{if } r \ll R_V \\ 1 & \text{if } r \gg R_V \end{cases} \tag{3.12}$$

$$\left|\frac{F_{\Pi_-}}{F_g}\right| \sim \frac{\Pi_-/M_4}{M/(M_4^2 r^2)} \sim \begin{cases} \left(\frac{r}{R_V}\right)^{3/2} & \text{if } r \ll R_V \\ \left(\frac{r}{R_V}\right)^{3} & \text{if } r \gg R_V. \end{cases} \tag{3.13}$$

These are both screening solutions, since the fifth force is suppressed on scales much smaller than R_V, but comparable to gravity at large scales. This is the simplest manifestation of the Vainshtein mechanism. For $0 \ll r \ll R_V$, both solutions $\Pi_\pm$ decay as $\sim 1/\sqrt{r}$, while for $r \gg R_V$, $\Pi_+ \sim 1/r^2$ and $\Pi_- \sim r$ as shown in figure 3.1. Note that in theories for which the quartic galileon is present, such as in the decoupling limit of massive gravity studied in the next section, the suppression is even stronger: $F_\pi/F_g \sim (r/R_V)^2$ for $r \ll R_V$ [41].

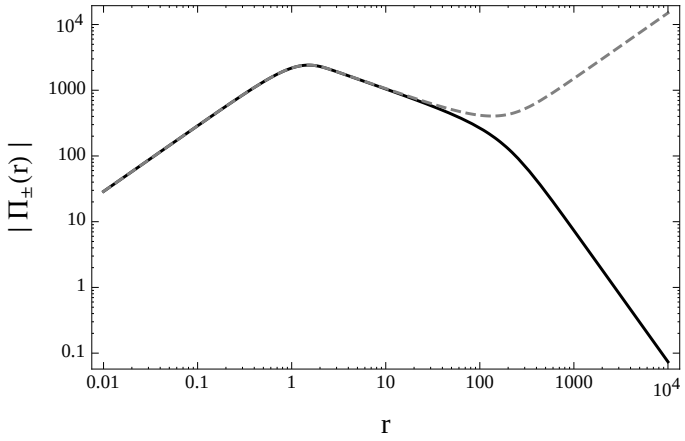

Figure 3.1: The two static screening solutions Π_+ (solid) and Π_- (dashed) of equation (3.10). These solutions correspond to a source radius of $R_0 = 1$ and a Vainshtein radius $R_V = 1000$, corresponding to a source density of $\rho \simeq 1.8 \times 10^8$. Both solutions grow as $\sim r$ inside of the source and decay as as $\sim 1/\sqrt{r}$ outside the source for $0 \ll r \ll R_V$. For very large distances $r \gg R_V$, Π_- diverges as $\sim r$, whereas $\Pi_+ \sim 1/r^2$.

The asymptotically decaying solution Π_+ gives rise to an asymptotically flat spacetime $g_{\mu\nu} \propto \eta_{\mu\nu}$, whereas Π_- sources a spacetime with cosmological asymptotics $g_{\mu\nu} = (1 - \frac{3}{4}\frac{\Lambda_3^3}{M_4} r^2)\eta_{\mu\nu}$. Note that this solution is distinct from the self-accelerating solution in DGP which gives rise to a maximally symmetric de Sitter spacetime. The self-accelerating solution has no sources and a scalar field configuration of $\pi = -(1/2)\Lambda_3^3 x_\mu x^\mu$, leading to a spacetime of the form $g_{\mu\nu} = (1 - \frac{1}{2}\frac{\Lambda_3^3}{M_4} x_\mu x^\mu)\eta_{\mu\nu}$ which is locally de Sitter (for times and physical distances much smaller than $L_{\rm DGP} \equiv M_4^2/M_5^3$) [129].

These static solutions of the DGP theory are well known, but very little is known about how equations of the type (3.9) behave in the fully dynamical regime. We attempt to address

this here by considering the linear and non-linear stability of the screening solutions, solving the full time-dependent equation of motion for the DGP scalar field numerically.

3.3.2 Dynamical stability analysis

Stability of the screening solutions to small perturbations is a fundamental condition for their physical relevance. The analysis of fluctuations around the static screening solutions reveals that Π_+ is stable to small high-frequency fluctuations, but Π_- is not (see also Appendix B). To see this, consider small perturbations $\delta\pi(r,t)$ about the screening solution, which have the action:

$$S_{\delta\pi} = \int d^4x \left[\frac{1}{2} Z^{\mu\nu}\partial_\mu\delta\pi\partial_\nu\delta\pi + \frac{1}{2M_4}\delta\pi T\right], \tag{3.14}$$

where the kinetic coefficients are given by the effective metric components

$$\begin{aligned} Z^{tt} = -\left(6 + 8\frac{\pi'}{r} + 4\pi''\right), \qquad Z^{rr} = 6 + 8\frac{\pi'}{r}, \\ r^2 Z^{\theta\theta} = r^2\sin^2\theta Z^{\phi\phi} = 6 + 4\frac{\pi'}{r} + 4\pi''. \end{aligned} \tag{3.15}$$

Therefore, although the field lives in flat space, it propagates in the effective metric $Z^{\mu\nu}$. For the asymptotically flat solution, the matrix $Z^{\mu\nu}$ has the signature of Minkowski spacetime. For the asymptotically non-flat solution, the kinetic terms switch sign in relation to other fluctuating degrees of freedom in Minkowski space. This "wrong" sign in the kinetic term indicates a ghost, first discussed in Ref. [128]. Dynamically, ghosts cause problems when there is an interaction with another field whose kinetic term takes the opposite sign. In the decoupling limit with fixed sources, the scalar field evolves independently, and we therefore do not expect to see any instabilities in either branch of solutions in what follows.

Fluctuations about the screening solutions can travel superluminally. Figure 3.2 shows the sound speed profiles, calculated as $c_s = \sqrt{-Z^{rr}/Z^{tt}}$, for both branches of screening solution. In both cases, there is superluminal propagation inside and outside of the source. Note that for large r, fluctuations around Π_+ travel luminally, while fluctuations around Π_- are subluminal and approach $1/\sqrt{2}$.

Although DGP has fallen out of favour due to the presence of ghost instabilities and superluminal propagation speeds [192, 193], both the ghostly and asymptotically flat branches give us insight into how solutions of more complicated, and theoretically consistent, cases of the general theory eq. (3.2) behave.

We have seen that physically appealing *static* profiles do exist as simple solutions to the equations of motion. Do these equilibrium configurations actually form from generic initial

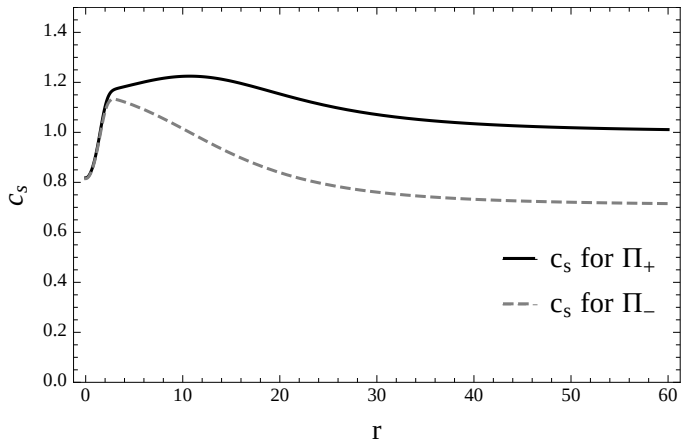

Figure 3.2: The sound speed profiles corresponding to fluctuations on top of the screening solutions, shown here for $R_0 = 1$ and $\rho = 10^5$, giving $R_V \sim 80$. Superluminal propagation $c_s > 1$ can occur in both branches inside and outside of the source. For large r, $c_s \to 1$ for the positive branch, while $c_s \to 1/\sqrt{2}$ for the negative branch.

data, or is there an unstable or ill-defined evolution? In other words, what possible initial configurations of $\pi(r, t = 0)$ – if any – would dynamically evolve to the screening solutions? Although a complete classification of the possible initial data is not possible, we focus on two representative cases: a spherical shell collapsing on the screening solution, and evolution from vacuum.

Code description

The Vainshtein screening solutions arise from the non-linear self-interactions of the π field. The solutions of interest are therefore fully in the non-linear regime of the equation of motion, eq. (3.9), where analytic solutions beyond the perturbative regime are difficult to obtain. Therefore, in order to fully explore the stability of screening solutions and investigate the general properties of time evolution, a numerical treatment is necessary.

We evolve eq. (3.9) numerically using the method of lines. Spatial derivatives were discretized on a fixed grid size with typical resolutions $dr = 1/25$, $dr = 1/50$ and $dr = 1/100$ using a second order differencing scheme. We explicitly checked that our results do not depend on the resolution. The resulting system of ordinary differential equations were then integrated using a fourth order explicit Runge-Kutta method. The stability of the evolution scheme relied on using stencils for both first and second derivatives; a fully first-order formulation of eq. (3.9) (in both spatial and time derivatives) did not yield stable numerical evolution. The spatial grid was constructed on the finite interval $[r_{\rm min},\ r_{\rm max}]$. Our results are independent of the interval chosen as long as $r_{\rm min}$ is sufficiently small and $r_{\rm max}$ sufficiently large such that the

outer boundary remains causally disconnected from the region under study. When dealing with the positive branch of solutions, we imposed Neumann boundary conditions at both ends $\pi'(r = r_{\rm min}) = \pi'(r = r_{\rm max}) = 0$, where this boundary condition at the origin is required for the field to remain regular. However, for solutions with cosmological asymptotics, the outer boundary condition was adjusted to $\pi'(r) = \Pi_-(r)$ at $r = r_{\rm max}$, where $\Pi_-(r)$ is from equation (3.11). The convergence of the code with increasing resolution is as expected for a second order scheme, see Appendix A. As an additional test of the code, many of our results were reproduced using the MATHEMATICA software.

Incoming spherical wave packet

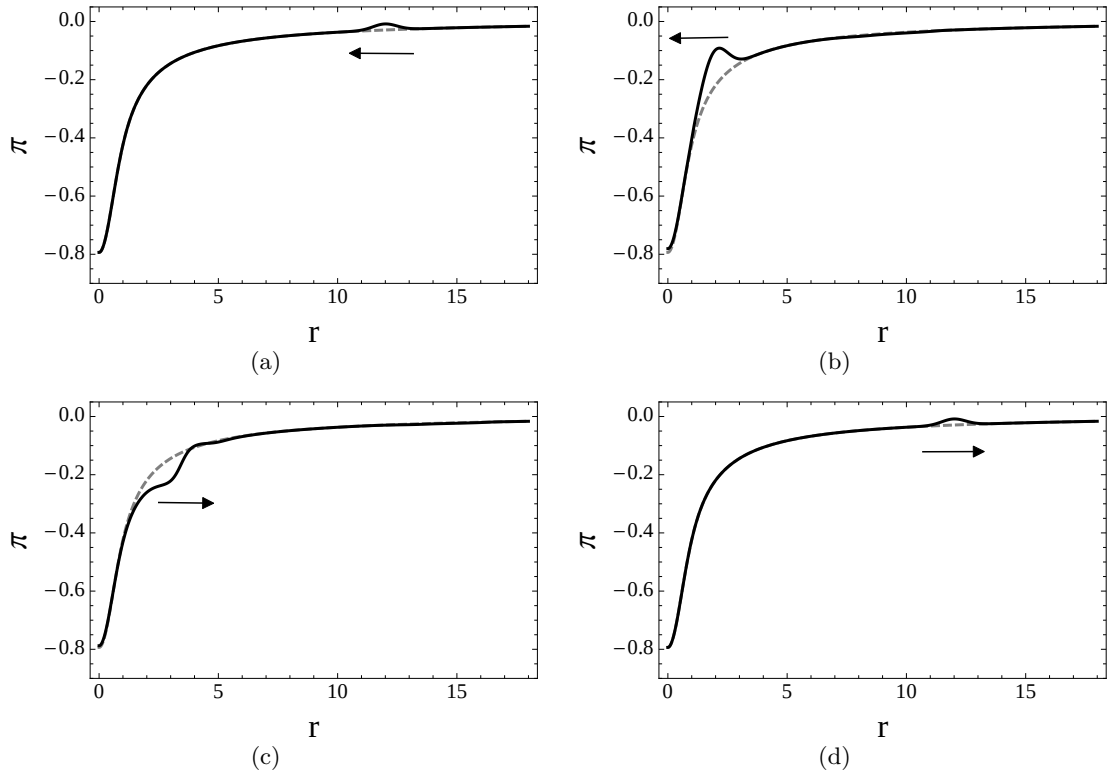

Figure 3.3: The evolution of $\pi(r,t)$ (solid) for an initial condition of the type (3.16) with $A = 0.25$, $\sigma = 0.5$, $r_w = 12$. The screening solution Π_+ is characterized by $\rho = 100$, $R_0 = 0.5$, $R_V \simeq 4$. From left to right, top to bottom are snapshots taken at $t = 0$, 10, 15, 23. The initial gaussian fluctuation is seen to propagate away from the domain at close to the speed of light, leaving behind the static screening solution at very late times.

The first family of initial data we consider are spherical shells collapsing on the static screening solutions (3.11). The source in the equation of motion is given by eq. (3.7), with variable ρ and R_0. To obtain a screening solution we must ensure that there is a significant hierarchy between R_0 and R_V. For an object like the sun, with $\rho \simeq 10^{26}$, we have $R_V \sim 10^9 R_0$. Resolving such a hierarchy of scales would be computationally intractable with our fixed grid

code; we must consider sources with a much more modest hierarchy. We use $R_0 = .5$ and $\rho = 100$, which gives a Vainshtein radius of $R_V = \sqrt{\pi}\rho^{1/3}R_0 \sim 8R_0$.

The initial data is given by:

$$\pi(r,0) = \pi_{\pm}(r) + \frac{Ar^2}{(r^2+\epsilon^2)^{3/2}} \exp\left(-\frac{(r-r_w)^2}{2\sigma^2}\right),$$
$$\dot{\pi}(r,0) = -\frac{Ar^2(r-r_w)}{\sigma^2(r^2+\epsilon^2)^{3/2}} \exp\left(-\frac{(r-r_w)^2}{2\sigma^2}\right). \tag{3.16}$$

where A and σ parametrize the amplitude and width of the shell, localized at r_w. The regulator $\epsilon \ll 1$ is to ensure that the field is well defined at $r = 0$. Note that this wave packet is purely in-going.

An example of the evolution is shown in figure 3.3, where the wavepacket is characterized by $A = 0.25$, $\sigma = 0.5$, $r_w = 12$. The incoming wavepacket perturbs the screening solution π_+, and then dissipates out of the computational domain leaving behind the original screening solution. Similar behavior was observed when perturbing the negative branch π_-. Exploring a wide range of A, σ, we find that the screening solution is stable to a range of nonlinear perturbations, although as we explain in more detail in section 3.4, large perturbations are problematic.

From vacuum to screening solutions

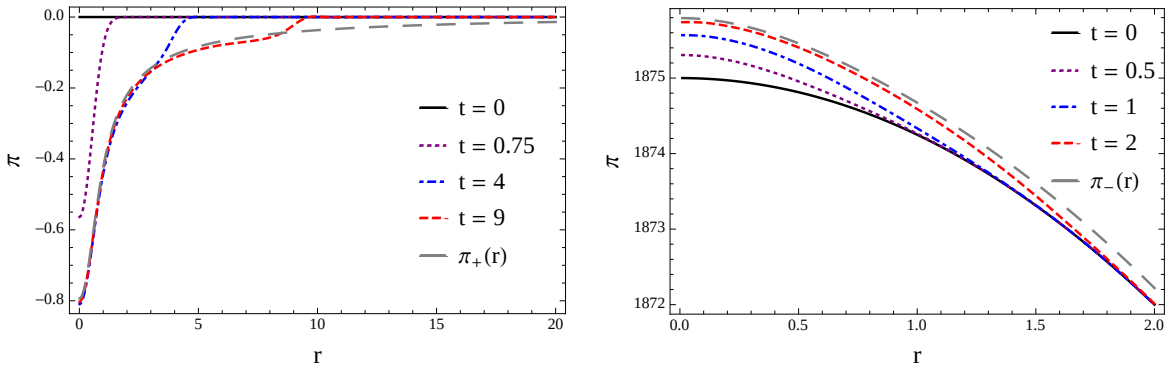

Figure 3.4: Left panel: The evolution of $\pi(r,t)$ with $\pi(r,0) = 0$ and $\rho = 100, R_0 = 0.5$, resulting in a Vainshtein radius $R_V \simeq 4$. At late times the solution approaches the static, asymptotically flat screening solution $\pi_+(r)$ corresponding to the same parameters.
Right: The evolution of $\pi(r,0) = -\frac{3}{4}(r^2 - r_0^2)$ for $\rho = 100$, $R_0 = 0.5$, $r_0 = 50$, $R_V \simeq 4$. The evolution drives the system towards the screening solution π_- at late times.

The second class of initial conditions we consider are the two vacuum solutions to eq. (3.9):

$$\pi(r,0) = 0, \qquad \pi(r,0) = -\frac{3}{4}(r^2 - r_0^2), \tag{3.17}$$

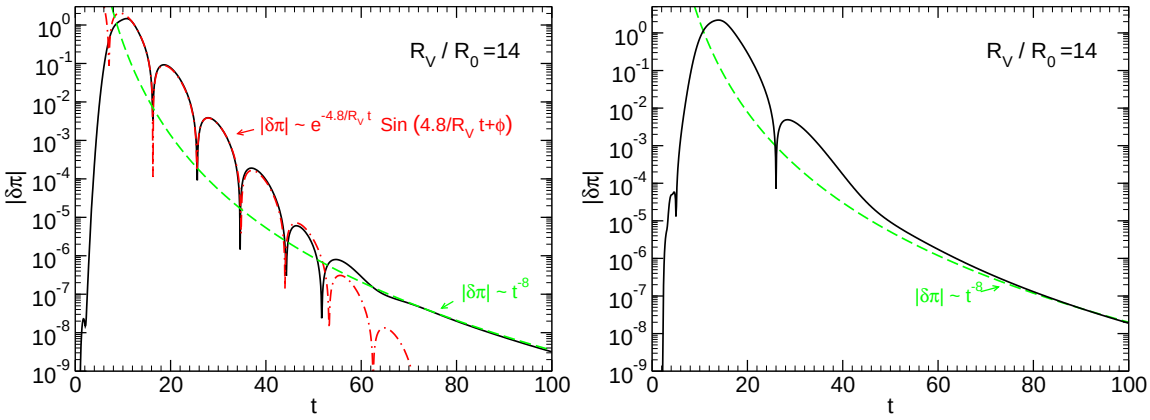

Figure 3.5: Left panel: Linearized time-evolution of a Gaussian wavepacket in the background of Π_+ with amplitude $A = 3$, width $\sigma = 1$ and localized at $r_w = 10$. The radius of the source is at $R_0 = 1$. The waves are extracted at $r = 2$. The intermediate-time behavior consists on an exponentially damped sinusoid –called quasinormal mode – and the late-time behavior is described by a power-law falloff of the field $\delta\pi \sim t^{-8}$. The observed behavior is in agreement with a frequency-domain numerical and analytical calculation; see text for further details.
Right panel: Same in the background of Π_-. The time-domain profile suggests that in this case the quality factor of the fundamental quasinormal mode ω_R/ω_I is smaller than in the background of Π_+, i.e. $\omega_R/\omega_I < 1$.

where r_0 is a normalization parameter. In the presence of a source, the vacuum initial conditions will necessarily evolve. A priori there are a number of possible endpoints to this evolution, but the expectation is that the static screening solutions $\pi_\pm$ are reached at late times. In order to determine the endpoint of evolution from both vacuum initial conditions, we perform numerical evolution in the presence of sources with varying R_V and R_0.

Examples of time evolution from vacuum initial conditions in the presence of a source with $\rho = 100$, $R_0 = 0.5$ are shown in figure 3.4. In these examples, the π_+ screening solution is reached from $\pi(r, 0) = 0$ and the π_- solution is reached from $\pi(r, 0) = -\frac{3}{4}(r^2 - r_0^2)$. The initial condition was set to match the screening solution at $r = r_0 = 50$, so the evolution is most visible for small r which is why we only plot up to $r = 2$. The fact that the expected screening solutions are reached as the endpoint of evolution from vacuum initial conditions displays that the screening solutions are quite robust to large perturbations.

3.3.3 Quasinormal modes and tails of screening solutions

The previous subsections focused on completely nonlinear evolution, and suggest that when perturbed, a static screening solution behaves as a coherent object: it vibrates and eventually settles down to the original static solution. To understand this behavior more thoroughly, and to understand generic small fluctuations of the screening solutions (3.11), we focus now on

linearized fluctuations, considering generic perturbations of the form

$$\pi(t, r, \vartheta, \varphi) = \pi_{\pm}(r) + \epsilon\, \delta\pi(t, r) Y_{lm}(\vartheta, \varphi)\,, \tag{3.18}$$

where $\pi_{\pm}$ is the static solution given by (3.11), $Y_{lm}(\vartheta, \varphi)$ are the usual spherical harmonics and ϵ is a small bookkeeping parameter. Inserting (3.18) in equation (3.8) and linearizing in ϵ we find the equation for $\delta\pi$,

$$Z^{tt}\ddot{\delta\pi} + Z^{rr}\delta\pi'' + 2rZ^{\theta\theta}\delta\pi' - l(l+1)Z^{\theta\theta}\delta\pi = 0\,, \tag{3.19}$$

where the coefficients Z^{tt}, Z^{rr} and $Z^{\theta\theta}$ are those in equation (3.15).

We evolved eq. (3.19) in time considering an initial Gaussian wave-packet $\dot{\delta\pi} = \frac{A}{r} e^{(r-r_w)^2/2\sigma^2}$. A typical waveform is shown in figure 3.5 for $R_0 = 1$ and $\rho = 500$. This plot shows the value of the field as a function of time at a specific position r, and has the same form regardless of the position. The waveform consists of three stages, familiar in the context of wave propagation in curved spacetimes [194, 195]: a prompt response at very early times, which depends on the details of the initial data and is the analogue of on-light-cone propagation in flat spacetime; at intermediate times the signal consists of a series of exponentially damped sinusoids, termed quasinormal modes [194] which correspond to the "characteristic modes" of the vibrating object. In this case the vibrating object is the static screening solution, and the fluctuations are damped because the system is intrinsically dissipative: energy flows to infinity. This stage is universal and independent of the details of the initial conditions. Finally, a power-law tail sets in at very late times caused by backscattering off the scalar profile (in complete analogy with backscattering due to spacetime curvature [196, 197]). Comparison against the full nonlinear evolution confirmed this typical behavior.

Quasinormal modes

To quantify the three stages of evolution, it is useful to recast the evolution equation as a Schrodinger-type equation in the frequency domain. Defining

$$\psi(t, r) = \delta\pi(t, r) \frac{r}{\sqrt{2}} \left(-Z^{tt} Z^{rr}\right)^{1/4}\,, \tag{3.20}$$

and a new coordinate r_* by

$$\frac{dr}{dr_*} = f(r) \equiv \sqrt{-\frac{Z^{rr}}{Z^{tt}}} = \sqrt{\frac{6 + 8\Pi/r}{6 + 8\Pi/r + 4\Pi'}}\,. \tag{3.21}$$

we can rewrite eq. (3.19) as a wave equation of the form

$$\left[\partial_{r_*}^2 - \partial_t^2 - V(r)\right] \psi(t, r_*) = 0\,, \tag{3.22}$$

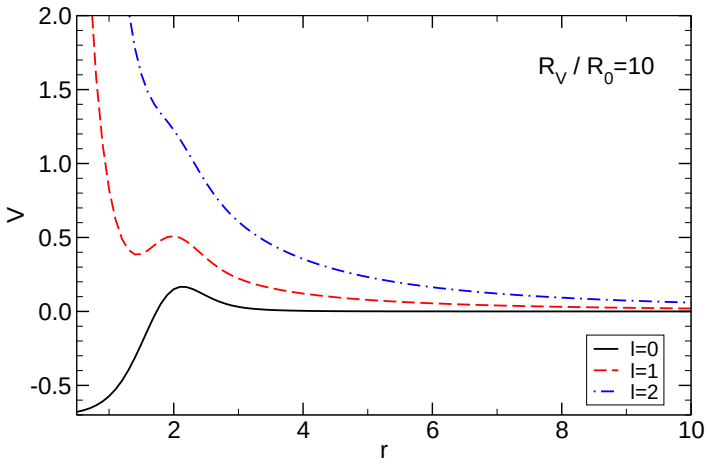

Figure 3.6: Effective potential in the background of Π_+ for $R_0 = 1$, $R_V = 10$ and different multipoles l. In the background of Π_- the potential is qualitatively similar.

where the effective potential V is given by

$$V = -\frac{r^2 Z^{\theta\theta}}{Z^{tt}} \frac{l(l+1)}{r^2} + \frac{4Z_2 + 3f^4(Z^{rr}-2)(Z^{rr}+6)}{4r^2 f^2 Z_2} + \frac{f^2 Z_2 (r^2 (f')^2 - 8) - 8rff'Z_2 - 2r^2 f^3 Z_2 f''}{4r^2 f^2 Z_2} \,. \tag{3.23}$$

Here, primes stand for radial derivatives and we defined $Z_2 \equiv (Z^{rr})^2$. The effective potential $V(r)$ in the background of Π_+ is plotted in figure 3.6 for the monopole, dipole and quadrupole components ($l = 0, 1, 2$). The shape is familiar from studies of wave quasinormal modes and tails around black holes and neutron stars [198, 194, 199]. The local maximum indicates that the effective metric $Z^{\mu\nu}$ allows for unstable null circular geodesics, with the instability timescale dictating the lifetime of fluctuations. In analogy with the gravitational cases, we expect the screening solutions to support quasinormal modes. These modes can be understood as the scalar modes of vibration of a coherent object. Unlike normal modes, they decay in time due to dissipation, where in this case dissipation occurs due to the leakage of energy to infinity (see Refs. [198, 194, 200] for reviews).

To perform a quantitative analysis, it is convenient to go to fourier space:

$$(\delta\pi(r,t), \psi(t,r)) = (\Delta\pi(r,\omega), \Psi(r,\omega))\, e^{-i\omega t} \,. \tag{3.24}$$

The equation of motion (3.19) and (3.22) are written as

$$Z^{rr}\Delta\pi'' + 2rZ^{\theta\theta}\Delta\pi' - \left(\omega^2 Z^{tt} + l(l+1)Z^{\theta\theta}\right)\Delta\pi = 0 \,, \tag{3.25}$$

$$\left[\partial_{r_*}^2 + \omega^2 - V(r)\right]\Psi(r_*,\omega) = 0 \,. \tag{3.26}$$

At the origin the equation of motion admits the behavior

$$\Delta\pi(r,\omega), \Psi(r,\omega)/r \sim A_1 r^l + A_2 r^{-(l+1)} \,. \tag{3.27}$$

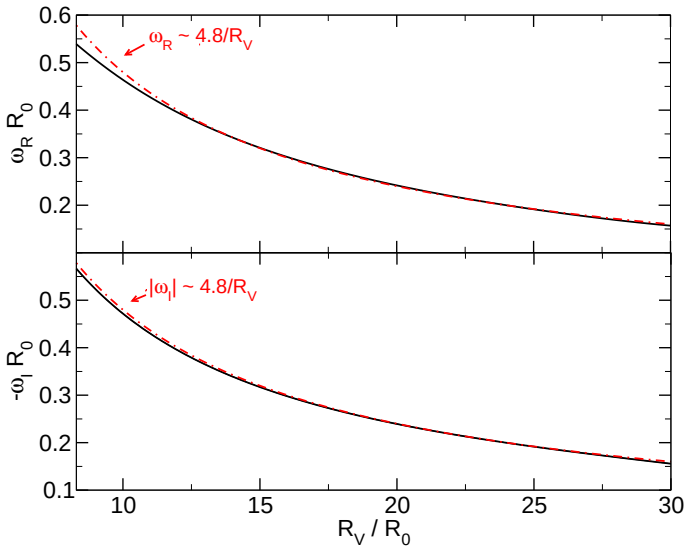

Figure 3.7: Fundamental quasinormal modes of the scalar field in the background Π_+. The full lines correspond to the numerical results, whereas the dashed lines show the analytic approximation at low frequencies. The top and bottom panels show the real part, $\omega_R R_0$, and the imaginary part, $\omega_I R_0$, of the mode as a function of the Vainshtein radius R_V/R_0.

Regularity of the field and its derivatives requires that $A_2 = 0$. At infinity one has

$$r\Delta\pi(r,\omega), \Psi(r,\omega) \sim B_1 e^{ik_\infty r} + B_2 e^{-ik_\infty r}\,, \tag{3.28}$$

where $k_\infty = \omega$ in the background of Π_+ and $k_\infty = \sqrt{2}\omega$ in the background of Π_-. Requiring that the system is otherwise isolated is tantamount to demanding Sommerfeld outgoing boundary conditions, $B_2 = 0$.

With the above two boundary conditions, Eqs. (3.25) or (3.26) define an eigenvalue problem for the (generically complex) quasinormal frequency $\omega = \omega_R + i\omega_I$. To compute the eigenfrequencies we use a direct integration approach described in Refs. [201, 202]. Notebooks are available online.[1] We integrate from each of the boundaries towards a matching point where the wavefunction and its radial derivative are required to be continuous. For generic ω the continuity conditions are not satisfied, unless ω is an eigenfrequency. One can then find the eigenfrequencies using a standard shooting method. The eigenfrequencies are typically ordered by increasing (absolute value of) imaginary part, the fundamental mode being the largest and longest-lived. The fundamental overtone for $l = 0$ is shown in figure 3.7 as a function of R_V/R_0, and is well approximated by

$$\omega_R \sim \frac{4.8}{R_V}\,, \tag{3.29}$$

$$\omega_I \sim -\frac{4.8}{R_V}\,. \tag{3.30}$$

[1] `https://blackholes.tecnico.ulisboa.pt/?page=Files`

This scaling is generic and does not depend on the details of the source. The frequency domain and time domain analysis agree extremely well with each other and with the non-linear results, as summarized in figure 3.5. In the background of Π_- the time-domain profile suggests that the quality factor ω_R/ω_I is smaller than in the background of Π_+, that is $\omega_R/\omega_I < 1$. The method used in the frequency-domain works well when $\omega_R \gtrsim \omega_I$, thus in the background of Π_- we have not been able to compute the quasinormal frequencies accurately. Nevertheless we expect the scaling to follow closely Eqs. (3.29) and (3.30).

These results are very general; perturbed screening solutions will ring, and the response will be dominated by its lowest quasinormal modes. In the present setting, gravitational degrees of freedom of the source are frozen. Once they are allowed dynamics, a second quasinormal mode stage will appear, corresponding to the oscillation of the source itself [195]. Source dynamics are presumably already well understood in GR, the ringdown stage we described is new and can be assigned entirely to the large-scale scalar field screening solution. The time scale of the ring-down is proportional to the Vainshtein radius and we explicitly checked that this scaling is independent on the source functional form, showing that the static solutions behave as large-scale objects localized at R_V.

Late-time power-law tails

A thorough study of the late-time behavior of equations of the form (3.26) was performed in Ref. [196, 197]. A late-time power-law tail of the form $\psi \sim t^{-\beta}$ is caused by back-scattering off the (effective) spacetime curvature *at large distances* (mathematically this is due to a branch cut in the Green's function) and has the form

$$\lim_{t\to\infty} \psi(r_*, t) = t^{-2l-\alpha}\,, \quad \text{for } \lim_{r_*\to\infty} V = \frac{l(l+1)}{r_*^2} + \frac{K}{r_*^\alpha}\,, \tag{3.31}$$

with l an integer and K is a constant that depends on l and $M(r\to\infty)$. In the background of $\Pi_\pm$ the effective potential (3.23) has the large distance asymptotic behavior

$$V \sim \frac{l(l+1)}{r_*^2} + \frac{K}{r_*^8}\,. \tag{3.32}$$

Thus, spherically symmetric fluctuations ($l = 0$) are expected to decay as t^{-8} at very late times for both the Π_+ and Π_- background solutions.[2] Such decay is consistent with our findings in figure 3.5. For higher multipoles, the above analysis predicts a decay $\psi(r_*, t) \sim t^{-2l-8}$ at late times.

Note that the features we described are generic to any kind of scalar field with non-linear

[2]For the special case of static initial data the power changes to t^{-9}.

kinetic terms. In particular, our analysis imply that perturbations of galileons [125] will display similar late-time behavior.

3.4 Cauchy Breakdown

We have shown in the previous sections that the static screening solutions are stable against a variety of fluctuations. In this section we want to give a quantitative measure of how robust the solutions are against *large* perturbations.

In theories described by eq. (3.2), scalar fluctuations propagate in an effective metric $Z^{\mu\nu}$. As shown in Appendix B, for the evolution problem to be well posed (i.e., that the solution is unique and depends continuously on the initial data), the initial data must be such that the effective metric $Z^{\mu\nu}$ has a Lorentzian signature, e.g., $\det(Z^{\mu\nu}) < 0$ everywhere in space, and surfaces of constant time are required to be spacelike, $Z^{tt} < 0$.

However, even within this restricted class of initial conditions, problems can still arise because of the non-linearity of the equations. Since the spacetime metric is in general different from the effective metric $Z^{\mu\nu}$, and it is possible that $Z^{tt} \to 0$ in the absence of any other pathologies like singularities or horizon formation (in fact we will mostly deal with a flat spacetime metric). For DGP the relevant components of the effective metric for a time-dependent background are given by

$$\begin{aligned} Z^{tt} &= -\left(6 + 8\frac{\pi'}{r} + 4\pi''\right), \quad Z^{rr} = 6 + 8\frac{\pi'}{r} - 4\ddot{\pi}\,, \\ Z^{tr} &= 4\dot{\pi}', \quad r^2 Z^{\theta\theta} = 6 + 4\frac{\pi'}{r} + 4\pi'' - 4\ddot{\pi}\,. \end{aligned} \tag{3.33}$$

If at any point in spacetime $Z^{tt} \to 0$, the Cauchy problem breaks down because the surfaces of constant time become null with respect to the effective metric, i.e., $Z^{\mu\nu}\partial_\mu t \partial_\nu t \to 0$. When this happens, the numerical evolution ceases to be possible past this point, and it is possible that the theory itself ceases to be well defined. Similar issues have been reported recently in the context of k-essence models [203, 204, 188]. We refer to this issue ($Z^{tt} = 0$) as Cauchy breakdown, following earlier nomenclature [188]. Substituting the static screening solutions into Z^{tt} gives that $Z^{tt} < -6$ for Π_- and $Z^{tt} > 12$ for Π_+, so Cauchy breakdown is not an issue initially. However, if we perturb the static screening solution, Z^{tt} could pass through zero at a finite r and t.

Besides the issue of Cauchy breakdown, we might also expect that in some situations, regions can form where $Z^{rr} \to 0$, $Z^{\theta\theta} \to 0$ or any of the eigenvalues of the matrix $Z^{\mu\nu}$ given by eq. (B.10) cross zero. For time-dependent backgrounds we have $Z^{tr} \neq 0$, so regions where

$Z^{rr} \to 0$ are not physical singularities but are rather regions where a sound horizon forms. This is a typical feature of non-linear fields with non-linear kinetic terms [205, 206, 203, 204, 188] and it is to be expected in regions where the propagation speed of the fluctuations is much smaller than the propagation speed of the background. However, as discussed in Appendix B, regions where $Z^{\theta\theta}$ (or any other eigenvalue of $Z^{\mu\nu}$) switch sign, are prone to instabilities. The timescale of these instabilities generically scales with $\tau \sim \Lambda_3^{-1}$. In a dynamical setup local instabilities can arise for a small amount of time, $t_{\rm din}$, and in small regions in space. These instabilities are not necessarily catastrophic as long as the instability timescale $\tau \gtrsim t_{\rm din}$. Furthermore, looking at (3.33) we expect that if any unstable region forms for long times, the development of this instability will make derivatives of the fields grow and will likely be followed by Cauchy breakdown.

Cauchy breakdown occurs in a wide variety of scenarios, which we can study using the numerical methods described in section 3.3.2. For the class of initial data described in eq. (3.16), Cauchy breakdown occurs at fixed σ for a sufficiently large amplitude A. Using this set of initial data for the Π_+ branch, neither sound horizon regions nor unstable regions form. On the other hand for Π_-, sound horizons and unstable regions can form for a finite time before Cauchy breakdown. If the fluctuations are sufficiently small these regions eventually disappear when the wave dissipates to infinity with a timescale smaller than the instability timescale. But if the fluctuations are sufficiently large, following the onset of these instabilities, Cauchy breakdown will always occur.

In figure 3.8, we show the evolution of Z^{tt} for a wave packet with $A = 1$, $\sigma = 0.5$, $r_w = 12$; evolution cannot proceed past $t \sim 9.8$, where $Z^{tt} \sim 0$.

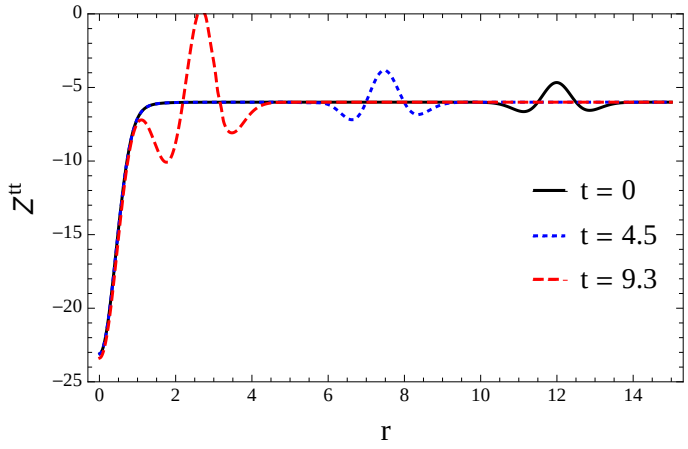

Figure 3.8: The evolution of Z^{tt} given in (3.15) to Cauchy breakdown, with the initial condition (3.16) with $A = 1$, $\sigma = 0.5$, $r_w = 12$. The source is characterized by $\rho = 100$, $R_0 = 0.5$, $R_V \approx 4$.

To understand the set of initial conditions described by eq. (3.16) that lead to well-defined time-evolution, we have performed an extensive search for Cauchy-breakdown in the Π_+ branch (results are qualitatively similar for the Π_- branch). Our results are summarized in figure 3.9. In the left panel, we fix the source properties to be $R_0 = 1$ and $\rho = 50, 200$ and study wavepackets with varying width. We also compare the results to the case where there is no source. The algorithm we used to find the critical amplitudes at which Cauchy-breakdown occurs is the following: starting from $\sigma_0 = 0.3$ and $A_{\rm init} = 0.01$, we increase A by steps of $\delta A = 0.01$ until we locate the critical amplitude $A_{\rm crit}$ at which $Z^{tt} \to 0$ is reached somewhere during the evolution (the time evolution terminates when this occurs). For amplitudes above this critical threshold, Cauchy breakdown occurs. When this happens we break the loop, starting a new loop for $\sigma_1 = \sigma_0 + \delta\sigma$ (we used $\delta\sigma = 0.05$), and using $A_{\rm crit}$ as the new $A_{\rm init}$ for this loop. We implemented this algorithm in the range $0.3 \leq \sigma \leq 1.5$. In the right panel, the width of the wave packet is fixed at $\sigma = 0.5$ and the size of the source fixed to be $R_0 = 1$. We then locate the critical amplitude at which Cauchy breakdown occurs for varying Vainshtein radius, or equivalently, varying central source densities ρ. We have explicitly checked that results for all simulations are independent of r_w, the initial pulse location, when $r_w \gtrsim R_V$.

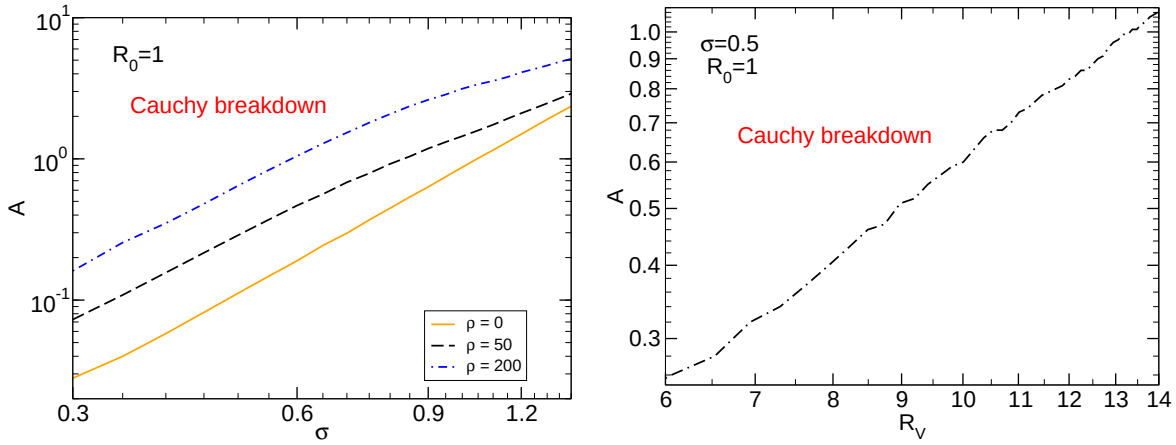

Figure 3.9: Log-Log plots of the region in parameter space of initial conditions for which Cauchy breakdown occurs. Breakdown occurs above each of the lines shown, which correspond to the critical amplitude $A_{\rm crit}$. Recall that the Vainshtein radius is defined by $R_V = \sqrt{\pi}\rho^{1/3}R_0$.

Examining the left panel of figure 3.9, there are several clear trends. In vacuum the critical amplitude is fairly well fit by a power law with $A_{\rm crit} \propto \sigma^3$. On the other hand, for both source densities, the critical amplitude is fitted by a broken power law with $A_{\rm crit} \propto \sigma^3$ at small widths and $A_{\rm crit} \propto \sigma^2$ at widths $\mathcal{O}(1)$ and larger. Comparing the two densities we have sampled, we can also conclude that higher density sources are more robust to Cauchy breakdown for a fixed perturbation amplitude. Since the perturbation is riding on a larger background screening solution, larger amplitudes are necessary to drive $Z_{tt} \to 0$. From the right panel of figure 3.9,

this increase in $A_{\rm crit}$ appears to scale roughly with $R_V^{3/2}$ (in terms of density, $A_{\rm crit}$ therefore scales like $\rho^{1/2}$).

For vacuum initial conditions eq. (3.17), we also find Cauchy breakdown, which is normally preceded by the formation of a sound horizon and an unstable region for both branches. To investigate the types of sources for which this occurs, we have simulated evolution in the presence of sources with radii between $0.1 \leq R_0 \leq 3$ and Vainshtein radii between $1 \leq R_V \leq 50$. For sufficiently small central densities, a wavepacket forms, taking the initial conditions to the final screening solution. Additional fluctuations are dissipated out of the computational domain, leaving the screening solution. For R_V, or equivalently ρ, larger than a critical threshold, the wavepacket overshoots the screening solution and the Z^{tt} factor to pass through zero causing Cauchy breakdown. We characterize the parameter space leading to Cauchy breakdown in figure 3.10. For $\pi(r,0) = 0$ we find that for values above $R_V/R_0 \sim 15.7$ there is Cauchy breakdown. We obtain a qualitatively similar result for quadratic vacuum initial conditions.

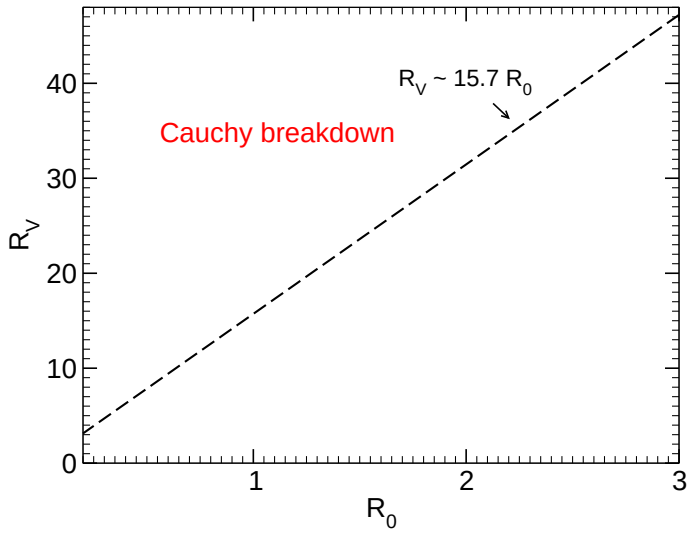

Figure 3.10: For the initial condition $\pi(r,0) = 0$, Cauchy breakdown occurs in the region above the curve. The trend shows that above $R_V/R_0 \sim 15.7$ there is Cauchy breakdown.

Based on these results, we see that the Vainshtein screening solutions in DGP are dynamically stable to a wide variety of perturbations. In all cases, as long as Cauchy breakdown is avoided, the screening solution is approached at late times. Sources with large central densities (and correspondingly large hierarchies between the source size and Vainshtein radius) are more robust to perturbations. A screened object is therefore most vulnerable when it is in a low density state – first starting to collapse. In general, the presence of sources makes the theory less susceptible to Cauchy breakdown, but there is nevertheless a restriction on initial data

that leads to well-posed evolution.

3.4.1 Coordinate invariance of Cauchy breakdown

Cauchy breakdown could be either a point where the theory breaks down [207] or an artificial problem due to the way we choose to slice the spacetime. In fact, locally, $Z^{\mu\nu}$ can always be brought to the Minkowski form by the appropriate coordinate transformation, as long as the hyperbolicity condition, $\det(Z^{\mu\nu}) < 0$, is met. However, since the matter fields evolve in the spacetime metric, we have to consider also the dynamics in the metric $\eta_{\mu\nu}$. To have a well-posed Cauchy problem we must have a common family of Cauchy surfaces with respect to $\eta_{\mu\nu}$ and the effective spacetime metric $Z^{-1}_{\mu\nu}$ [208], where $Z^{-1}_{\mu\nu}$ is the inverse of $Z^{\mu\nu}$, i.e., $Z^{-1}_{\mu\nu} Z^{\mu\nu} = \delta^{\nu}_{\mu}$. If $\det(Z^{\mu\nu}) \neq 0$, for spherically symmetric spacetimes we have

$$
\begin{aligned}
Z^{-1}_{tt} &= \tfrac{Z^{rr}}{Z^{tt}Z^{rr} - (Z^{tr})^2}\,, \qquad & Z^{-1}_{rr} &= \frac{Z^{tt}}{Z^{tt}Z^{rr} - (Z^{tr})^2}\,, \\
Z^{-1}_{tr} &= \tfrac{Z^{tr}}{(Z^{tr})^2 - Z^{tt}Z^{rr}}\,, \qquad & Z^{-1}_{\theta\theta} &= \frac{1}{Z^{\theta\theta}}\,.
\end{aligned}
\tag{3.34}
$$

Consider a general spacelike hypersurface, with respect to Minkowski spacetime, with unit normal n^{μ} such that $\eta_{\mu\nu} n^{\mu} n^{\nu} = -1$. Working in spherical symmetry, and only considering general coordinate transformations of r and t, the most general unit normal is:

$$
n^{\mu} = (A(r,t), \sqrt{A(r,t)^2 - 1}, 0, 0)\,, \tag{3.35}
$$

for an arbitrary spacetime function $|A| > 1$. We want to know if there is any choice of A which keeps the unit normal timelike with respect to the effective spacetime $Z^{-1}_{\mu\nu}$:

$$
Z^{-1}_{\mu\nu} n^{\mu} n^{\nu} < 0\,. \tag{3.36}
$$

Expanding out we obtain,

$$
Z^{-1}_{tt} A^2 + 2 Z^{-1}_{tr} A \sqrt{A^2 - 1} + Z^{-1}_{rr}(A^2 - 1) < 0\,. \tag{3.37}
$$

In particular, when Cauchy breakdown occurs, we have $Z^{-1}_{rr} = \frac{Z^{tt}}{Z^{tt}Z^{rr} - (Z^{tr})^2} = 0$, so the above inequality simplifies to:

$$
Z^{-1}_{tt} < -2 Z^{-1}_{tr} \sqrt{1 - A^{-2}}\,. \tag{3.38}
$$

In the limit of large A, we have:

$$
Z^{-1}_{tt} < -2 Z^{-1}_{tr}\,. \tag{3.39}
$$

In the limit of $A = 1 + \epsilon$ with $\epsilon \ll 1$, we have:

$$
Z^{-1}_{tt} < -\sqrt{2\epsilon}\, Z^{-1}_{tr} \tag{3.40}
$$

The stronger condition is the first one. Using eq. (3.34) we conclude that it is possible to find a common spacelike surface in both the flat and effective spacetimes only when:

$$Z^{rr} < 2Z^{tr}\,. \tag{3.41}$$

For all cases we studied we found that this is never satisfied when Cauchy breakdown occurs. To understand this consider the DGP model with the effective metric given by (3.33). Using $Z^{tt} = 0$ and (3.33), the condition (3.41) reads

$$-4(\pi'' + 2\dot{\pi}' + \ddot{\pi}) < 0\,. \tag{3.42}$$

Cauchy breakdown generically occurs inside regions where gradients become large and do not have a definite sign. For example, for large fluctuations of the background static solutions, they occur at the peak of the traveling wave packet (see figure 3.8). Approximating the wave packet by a Gaussian of the form (3.16), we see that second derivatives are all negative at the peak of the Gaussian. Thus eq. (3.42) is not satisfied there. This means that, for the cases we considered, Cauchy breakdown is a real physical problem and not simply an artificial coordinate singularity.

As a final remark, notice that Z^{tt} changes sign between Π_+ and Π_-, which is a result of the fact that $\Pi_+ > 0$ and $\Pi_- < 0$. This means that there is no way for time evolution to connect one branch of solutions to the other without going through a region where Cauchy breakdown occurs. In addition, it is impossible to construct a spacetime containing local regions of each branch. For example, one cannot match the negative branch on $r < r_*$ onto the asymptotically flat solution for $r > r_*$ since Z^{tt} would have to cross zero. An indication that such solutions do not exist (beyond the decoupling limit) has been presented in Ref. [209] (for similar considerations in dRGT massive gravity see Ref. [210]).

3.5 Collapsing and Exploding Sources

In this section we consider a dynamical source $T(r,t)$ in order to model astrophysical phenomena where the source undergoes gravitational collapse into a relativistic object (e.g. a neutron star) or explodes (e.g. as in a supernova). Our models for the source are not physical in the sense that there is no underlying model, but rather are intended to give qualitative information on the possible evolution of π.

The first example we consider is a *collapsing* source. For simplicity, we assume that the relative contribution from pressure and density change in time, but the source radius does not.

Our model for the energy momentum tensor is:

$$T_{00} = \rho \exp\left(-\frac{r^2}{R_0^2}\right), \qquad T_{xx} = T_{yy} = T_{zz} = \frac{\rho}{3}\exp\left(-\frac{r^2}{R_0^2}\right)\left[1 - e^{-t/\tau}\right]. \tag{3.43}$$

Thus, the source becomes relativistic, with an equation of state $p = \rho/3$ on a timescale τ. We begin with the field at rest in one of the screening solutions $\Pi_\pm$. For adiabatic collapse $\tau \to \infty$, the field evolves as the source collapses to reach the vacuum solutions described above (either $\pi = 0$ or $\pi \sim r^2$). However, a source which collapses instantaneously corresponds to a large perturbation around the vacuum solutions, which from the results of section 3.4, can be vulnerable to Cauchy breakdown. This suggests a critical collapse time constant, $\tau = \tau_c$, below which Cauchy breakdown occurs.

In figure 3.11, a sample evolution is plotted for a collapsing source defined by parameters $\rho = 2000$, $R_0 = 1$, $\tau = 1$. In this example, the perturbation created by the collapse causes Cauchy breakdown at $t = 5$. Increasing the value of τ allows for a well-defined evolution. By fixing $R_0 = 1$ and varying ρ, we found that as the density increased, the corresponding time constant increased as well, in a linear fashion. Specifically, for both branches of solution, we found that $\tau_c \sim 0.0005\rho$.

Before Cauchy breakdown, the collapse seems to be generically preceded by the formation of an unstable region near the origin, where $Z^{\theta\theta}$ changes sign. Nevertheless, we always observe that the dynamical evolution is eventually stopped by $Z^{tt} \to 0$ before eventual unstable modes have time to grow.

Now let's consider the opposite effect: an *exploding* source. We once again begin with an initially screened source, and then model the "explosion" as an outgoing spherical shell of dust travelling at the speed of light:

$$T_{\mu\nu} = \mathrm{diag}\left(\frac{\rho}{f(t)}\exp\left(-\frac{(r-t)^2}{R_0^2}\right), 0, 0, 0\right). \tag{3.44}$$

where $f(t)$ is defined so that the mass at infinity has the same constant value as previously, specifically, $M(r \to \infty) = \pi^{3/2}\rho R_0^3$ for all t. This implies that

$$f(t) = (1 + 2t^2/R_0^2)(1 + \mathrm{Erf}(t/R_0)) + (2t/(\sqrt{\pi}R_0))e^{-t^2/R_0^2}. \tag{3.45}$$

We stress that this is not a physical model for an exploding source, but rather a simple test of the response to dynamical sources.

Starting with a screening configuration, as the source explodes, the scalar field relaxes to its vacuum state, while emitting a wave packet that travels with the source as it moves off to infinity. To illustrate this phenomenon, we show the evolution of the field in figure 3.12 for the case of a source with $\rho = 1500$ and $R_0 = 1$. Once again, one can imagine that for a source

that is sufficiently dense, a sound horizon and an unstable region can form when the source explodes due to regions where fluctuations propagate subluminally. These will generically happen before $Z^{tt} \to 0$ and can eventually disappear if the source is not too dense. However if the source is too dense, the induced perturbation in the scalar field will be enough to drive Z^{tt} to zero, causing Cauchy breakdown. The right panel of figure 3.12 shows the evolution of the Cauchy breakdown factor corresponding to the scalar field shown in the right panel. For this example, Z^{tt} safely avoids crossing zero, so the evolution remains well-defined.

To analyze the possibility of Cauchy breakdown in more detail we calculated the critical density ρ_c for various source sizes R_0, so that Cauchy breakdown is inevitable for $\rho > \rho_c$. As the source size increases, the corresponding critical density increases as $\rho_c \sim 2000 R_0$ (consistent with both branches). Once again, very dense compact objects can cause problems for the evolution.

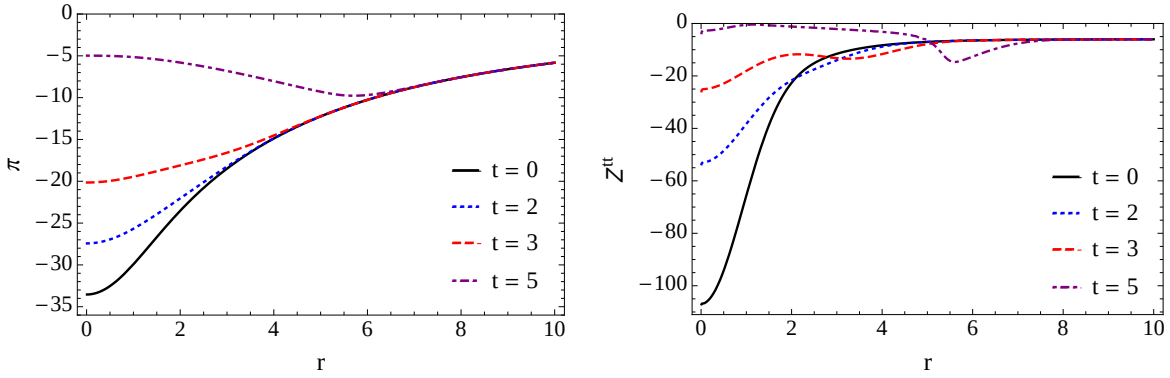

Figure 3.11: An example of evolving towards Cauchy breakdown for a collapsing source of the form (3.43) with $\rho = 2000$, $R_0 = 1$ and $\tau = 1$. Left panel: the evolution of $\pi(r,t)$ starting from $\pi_+(r)$. It is driven towards the vacuum $\pi_(r,t) \to 0$ as the source collapses (until Cauchy breakdown is reached).
Right: The corresponding factor Z^{tt} (3.33) that crosses zero at $t = 5$ resulting in Cauchy breakdown.

There are two relevant timescales for explosion or collapse: the crossing time of the source R_0 and the crossing time for the screening solution R_V. The longer timescale, R_V, sets the response time for the screening profile to changes in the source. In realistic scenarios, $R_0 \ll R_V$, and collapse or explosion will occur effectively instantaneously on timescales over which the screening solution can adjust. Therefore, we conjecture that Cauchy breakdown will be a problem for any realistic violent astrophysical process. However, to study breakdown in more detail, it is necessary to go beyond the decoupling limit and consider realistic dynamical sources.

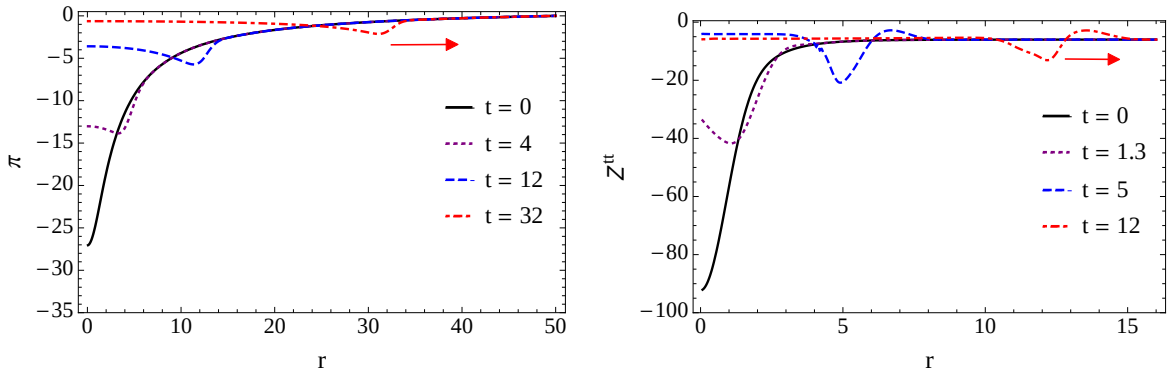

Figure 3.12: A sample evolution for an exploding source that evades Cauchy breakdown. The source has the form (3.44) with $\rho = 1500$ and $R_0 = 1$. Left panel: the evolution of $\pi(r,t)$ starting from $\pi_+(r)$. It is driven towards the vacuum $\pi_(r,t) \to 0$ as the source explodes, while a localized packet follows the travelling source off to infinity.
Right: The corresponding factor Z^{tt} (3.15). A perturbation is created that safely travels off to infinity without crossing zero, thus avoiding Cauchy breakdown.

3.6 Asymmetric screening solutions

In the previous sections we have been concerned about the linear and non-linear stability of the spherically symmetric screening solutions. We now wish to understand if these screening solutions can be generalized when we give up spherical symmetry.

Our numerical search for quasinormal modes in section 3.3.3 did not yield any zero-frequency mode. In other words, we were not able to find any regular, asymptotically flat static solution to the linearized equation of motion (3.19) (apart from the trivial solution for $l = 0$). This can be considered a simple version of a "no-hair" result for screening solutions: no static multipoles – other than the spherically symmetric monopole – are allowed to anchor onto spherically symmetric sources.

Does this result generalize for non-spherically-symmetric sources? Do the scalar multipoles anchor on higher source-multipoles? One possibility to study this issue is to look for nonlinear, asymmetric solutions. Given the structure of the equations of motion, such solutions are not trivial to find, although particular solutions can be built. Take for instance

$$\pi = \pi_1 \cos^2 \vartheta \,. \tag{3.46}$$

The field equations yield powers of $\cos\vartheta$ which can be matched to T order by order in powers of $\cos\vartheta$. Given a zeroth-order source function, the zeroth-order equation can be solved for π_1, and the remaining equations will then determine the source multipoles. For example, with the ansatz eq. 3.46 we find the following solution for the equations of motion (3.8), yielding a

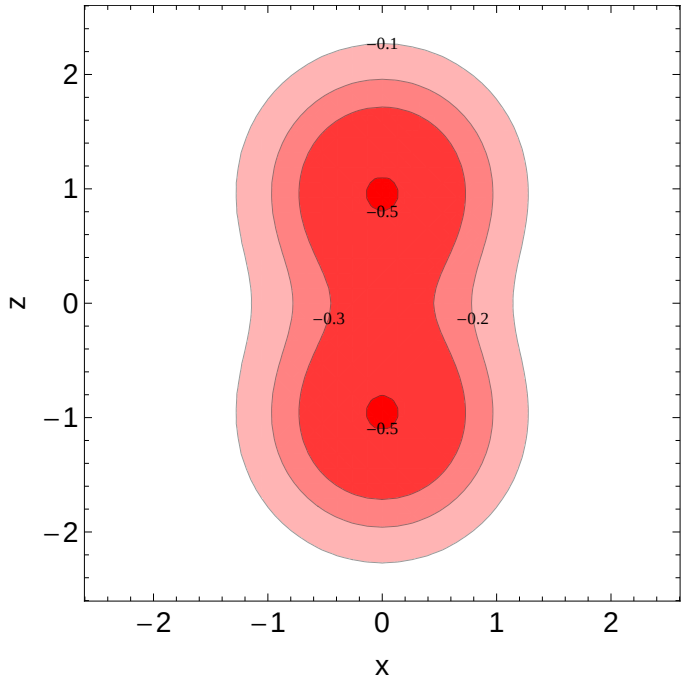

Figure 3.13: Contour plot of the $y = 0$ slice of density profile (3.51) describing two lumps of matter, here for $\rho = 1$, $z_0 = 0.9$, $R_0 = 1$.

quadrupolar static solution

$$\pi = -\frac{r^2}{24} T_0 \cos^2\vartheta \,, \tag{3.47}$$

$$T = T_0(r) + \frac{r}{36}\left(T_0'(108 - T_0 + 2rT_0')\right)\cos^2\vartheta - \frac{r}{36}\left(rT_0''(T_0 - 18)\right)\cos^2\vartheta + \frac{r}{72}\left(-T_0'(2T_0 + 13rT_0') + 2rT_0''(T_0 - rT_0')\right)\cos^4\vartheta \,. \tag{3.48}$$

This nonlinear solution represents a field strongly localized close to the source. However, for most quadrupolar source distributions $T(r, \theta)$, $-T$ is not positive definite implying that there are regions where $\rho < 3p$. Nonlinear solutions for higher multipoles can be found with the same scheme; they share similar properties.

A more robust method to look for asymmetric solutions builds on the nonlinear spherically symmetric solution for spherical sources (3.11). Realistic stellar – and other – sources are approximately spherically symmetric, and it is therefore appropriate to search for small deviations from spherical symmetry in both the source and the field. Specifically, we expand

$$T = T_0(r) + \epsilon \sum_{l=1} t_{lm}(r) Y_{lm}(\vartheta, \varphi) \,, \tag{3.49}$$

$$\pi(r) = \int^r \Pi_{\pm}(u) du + \epsilon \sum_{l=1} \delta\pi(r) Y_{lm}(\vartheta, \varphi) \,. \tag{3.50}$$

We defined $T_0 \equiv 2\sqrt{\pi} T_{00}$ and take this to be the dominant contribution. The components t_{lm} can be directly related to the more standard density moments $v_{lm} \propto t_{lm}(r) r^{l+2} dV$ where dV

is the volume element, in terms of which the gravitational potential multipole moments are usually expressed. Note that the dipole component vanishes for sources which are symmetric around the equator, while the quadrupole component is directly tied to the inertia tensor. For example, let's take two clumps of matter of the form (3.7), describing a deformed body,

$$T(r) = -\frac{\rho}{2}\left(e^{-(x^2+y^2+(z-z_0)^2)/R_0^2} + e^{-(x^2+y^2+(z+z_0)^2)/R_0^2}\right) . \tag{3.51}$$

This distribution represents two bodies localized at $\pm z_0$ on the z-axis, and is axially symmetric as shown in figure 3.13. For $z_0 = 0$ we recover the density distribution (3.7), so z_0 can be treated as an expansion parameter. All multipoles moments t_{lm} vanish for $m \neq 0$ (because the distribution is axially symmetric) and for odd l (because it has equatorial symmetry). For small z_0, the lowest multipoles are

$$T_0 = 2\sqrt{\pi}t_{00} = -\rho\, e^{-r^2/R_0^2} , \tag{3.52}$$

$$t_{20} = \frac{8}{3}\sqrt{\frac{\pi}{5}}\frac{r^2 z_0^2}{R_0^4}\, T_0 , \tag{3.53}$$

$$t_{40} = \frac{32\sqrt{\pi}}{315}\frac{r^4 z_0^4}{R_0^8}\, T_0 . \tag{3.54}$$

For a general source with multipoles $t_{lm}(r)$, the equation of motion for $\delta\pi$ is given by:

$$4r^2(3r+4\Pi_\pm)^2\delta\pi'' + 2\left(-12r\Pi_\pm - 8\Pi_\pm^2 + r^2(-18+T_0)\right)\left(l(l+1)\delta\pi - 2r\delta\pi'\right) = -r^3(3r+4\Pi_\pm)t_{lm}$$

This inhomogeneous ordinary differential equation can easily be integrated to yield solutions for $\delta\pi$. Solutions exist for any source and decay at large distances as r^{-l-1}.

Computing the components $t_{lm}(r)$ of the source eq. (3.51) around the Π_+ background, we integrate the equation above requiring regularity at the origin and vanishing field at infinity. This can be done using a standard shooting method using the constant A_1 of the expansion at the origin (3.27) as a shooting parameter. Some solutions are shown in figure 3.14 (results for Π_- are qualitatively similar). For very large densities the screening behavior of the different multipoles is apparent. For $R_0 \ll r \ll R_V$ the field decays as $r^{-l/2}$ indicating that in the Vainshtein regime higher multipoles have a stronger suppression than the monopole. The suppression of their contribution to the fifth force compared to the *multipoles* of the Newtonian gravitational force $F_g \sim r^{-(2+l)}$ are given by

$$\left|\frac{F_{l>0}}{F_g}\right| \sim \left(\frac{r}{R_V}\right)^{1+l/2} \quad \text{if } R_0 \ll r \ll R_V . \tag{3.55}$$

Finally, our results show that for $R_0 \ll r \ll R_V$ the field is proportional to $\sqrt{\rho} z_0^l$, whereas for very large distances it is proportional to ρz_0^l, for any multipole l. We conclude that tidal forces due to the scalar are subdominant to gravitational tidal forces inside the Vainshtein radius, and so the spherically symmetric approximation is, in general, a very good approximation to compute the fifth force around non-spherical stars.

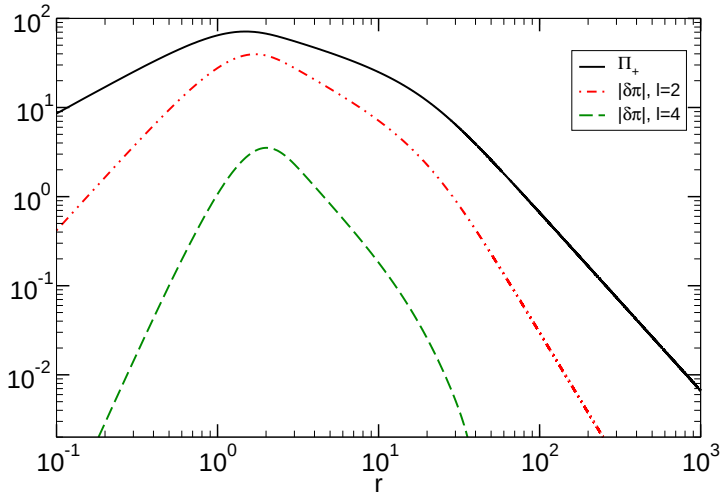

Figure 3.14: Hairy solutions for different multipoles l, here shown for $R_V = 100$, $z_0 = 0.9$, and compared to Π_+. For $R_0 \ll r \ll R_V$ the field decays as $\sqrt{\rho} z_0^l r^{-l/2}$ while for $r \gg R_V$ they decay as $\rho z_0^l r^{-(l+1)}$.

3.7 Screening in the Decoupling Limit of dRGT

3.7.1 dRGT equations of motion

Let us now make a comparison with ghost-free dRGT massive gravity which shares many features of the DGP model described above in the decoupling limit. Here, we consider a specific class of massive gravity in which one of its two free parameters is set to zero, making it possible to completely decouple the scalar and tensor modes. We'll see that a few features of massive gravity make the analysis different than DGP: a quartic galileon term, an extra free parameter α in the theory, and an extra non-minimal coupling to matter in the equations of motion of the form $\nabla_\mu \nabla_\nu \phi T^{\mu\nu}$. Strong constraints on this coupling can be found in Refs. [211, 212]. Also note that for a time-dependent field this coupling will change the Vainshtein radius [211]. However for small fluctuations around the screening solutions, this change is in general highly suppressed so we will keep working with the static quantity R_V.

The scalar mode that arises in massive gravity is essentially the longitudinal mode of the graviton, and in the decoupling limit (described in section 2.2.2) its dynamics can be solved for, independently of the other degrees of freedom.

Varying the decoupling limit dRGT Lagrangian (2.21) with respect to ϕ can be done using $\delta \mathcal{L}_n / \delta \phi = -2\mathcal{L}_{n+1}/(\partial\phi)^2$, so that the resulting scalar equation of motion is

$$\begin{aligned} 0 = &\frac{T}{M_4} - 2\frac{\alpha}{M_4 \Lambda_3^3} \nabla_\mu \nabla_\nu \phi T^{\mu\nu} + 3\Box\phi - 3\frac{\alpha}{\Lambda_3^3}\left[(\Box\phi)^2 - (\nabla_\mu\nabla_\nu\phi)(\nabla^\mu\nabla^\nu\phi)\right] \\ &+ \frac{\alpha^2}{\Lambda_3^6}\left[(\Box\phi)^3 - 3\Box\phi\,(\nabla^\mu\nabla^\nu\phi)(\nabla_\mu\nabla_\nu\phi) + 2\,(\nabla_\mu\nabla_\nu\phi)(\nabla^\nu\nabla_\gamma\phi)(\nabla^\gamma\nabla^\mu\phi)\right] \end{aligned} \tag{3.56}$$

Small perturbations around a time-dependent background propagate on the effective metric $Z^{\mu\nu}$, which can be computed perturbing eq. (3.56). For a spherically symmetric background the components of this metric are given by:

$$\begin{aligned}
Z^{tt}_{\rm MG} =& -3-\alpha\left(2T^{00}-6\phi''-12\frac{\phi'}{r}\right)-\alpha^2\left(\frac{6(\phi')^2}{r^2}+\frac{12\phi''\phi'}{r}\right), \qquad (3.57)\\
Z^{rr}_{\rm MG} =& 3+\alpha\left(-2T^{11}+6\ddot{\phi}-12\frac{\phi'}{r}\right)+\alpha^2\left(\frac{6(\phi')^2}{r^2}-\frac{12\ddot{\phi}\phi'}{r}\right),\\
Z^{tr}_{\rm MG} =& -6\alpha\dot{\phi}'+12\alpha^2\frac{\dot{\phi}'\phi'}{r},\\
r^2Z^{\theta\theta}_{\rm MG} =& 3+\alpha\left(-\tilde{T}^{22}-\tilde{T}^{33}+6\ddot{\phi}-6\phi''-6\frac{\phi'}{r}\right)+6\alpha^2\left(\frac{\phi''\phi'}{r}-\frac{\ddot{\phi}\phi'}{r}-\phi''\ddot{\phi}+(\dot{\phi}')^2\right),
\end{aligned}$$

where we defined $\tilde{T}^{22}\equiv T^{22}r^2$ and $\tilde{T}^{33}\equiv T^{33}r^2\sin^2\theta$ such that $\tilde{T}^{ii}$ are functions of r and t only.

Under the assumption of spherical symmetry $\phi=\phi(t,r)$, the equation of motion (3.56) for the longitudinal mode of the massive graviton in the decoupling limit for a non-relativistic source ($T_{0i}=T_{ij}=0$) is

$$\begin{aligned}
0 =& T-2\alpha T^{00}\ddot{\phi}-3\ddot{\phi}+3\phi''+\frac{6}{r}\phi'+6\alpha\left(\frac{2}{r}\ddot{\phi}\phi'-\frac{1}{r^2}(\phi')^2-(\dot{\phi}')^2+\ddot{\phi}\phi''-\frac{2}{r}\phi'\phi''\right)\\
&+6\alpha^2\left(\frac{2}{r}\phi'(\dot{\phi}')^2-\frac{1}{r^2}\ddot{\phi}(\phi')^2-\frac{2}{r}\ddot{\phi}\phi'\phi''+\frac{1}{r^2}\phi''(\phi')^2\right). \qquad (3.58)
\end{aligned}$$

Recall that we are imposing our choice of units: $M_4=\Lambda_3=1$. Further, we assume the same static source of density ρ, radius R_0 and mass M as given in (3.7). For the metric, the spherically symmetric ansatz is $h_{00}=a(r,t)$ and $h_{ij}=f(r,t)\delta_{ij}$. Once a solution for $\phi(r,t)$ is found, one can then find the metric functions using equations that result from variation of (2.21) with respect to $h_{\mu\nu}$:

$$f'=-\frac{M}{r^2}+\phi'(1-\alpha\frac{\phi'}{r}), \qquad (3.59)$$

$$a'=-\frac{M}{r^2}+r\ddot{\phi}-r\ddot{f}-\phi'-2\alpha\ddot{\phi}\phi'. \qquad (3.60)$$

Analysis of *static* screening solutions can be found in Refs.[174, 191]. Let us highlight some of their results. In this case, we can write (3.58) as a cubic (rather than quadratic as in DGP) polynomial in $\lambda=\phi'/r$:

$$3\lambda-6\alpha\lambda^2+2\alpha^2\lambda^3=\frac{M(r)}{4\pi r^3}. \qquad (3.61)$$

There are three solutions to (3.61), that we denote as λ_1, λ_2, λ_3. When the solutions are evaluated for small values of r outside of the source ($R_0<r\ll R_V$), it is clear that only λ_1 is real, and λ_2, λ_3 are imaginary in this regime. Therefore, we can take λ_1 to be our static

solution and disregard $\lambda_{2,3}$. The expression for λ_1 in this limit is valid for positive or negative values of α and can be written

$$\lambda_1(R_0 < r \ll R_V) \sim \frac{1}{\alpha} + \frac{R_V}{|\alpha|^{2/3} r} + \frac{r}{2|\alpha|^{4/3} R_V}\,. \tag{3.62}$$

For the sake of completeness, we give the full expression here

$$\begin{aligned}\lambda_1 =& \frac{1}{\alpha} + \frac{(2\pi)^{1/3} r}{\alpha\left(4\pi r^3 + \alpha M(r) + \sqrt{-16\pi^2 r^6 + \alpha M(r)\left(8\pi r^3 + \alpha M(r)\right)}\right)^{1/3}} \\ &+ \frac{\left(4\pi r^3 + \alpha M(r) + \sqrt{-16\pi^2 r^6 + \alpha M(r)\left(8\pi r^3 + \alpha M(r)\right)}\right)^{1/3}}{2\alpha(2\pi)^{1/3} r}\,.\end{aligned} \tag{3.63}$$

The interesting fact is that the asymptotic behavior of λ_1 is very different depending on the sign of α:

$$\lambda_1(r \to \infty) = \begin{cases} \frac{3+\sqrt{3}}{2\alpha} & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha < 0\,. \end{cases} \tag{3.64}$$

Plugging the asymptotics (3.64) into (3.59), (3.60), we see that when $\alpha < 0$ it is possible to have an asymptotically flat spacetime: $a \simeq f \simeq M/r$. However, when $\alpha > 0$, we obtain a non-trivial background with cosmological asymptotic behaviour: $a \simeq -r^2\lambda_1/2$, $f \simeq r^2\lambda_1(1-\alpha\lambda_1)/2$. This begs the question: what value can this free parameter α take? It was shown in [191] that $\alpha > 0$ is required to avoid a ghost instability, so let us examine this in more detail.

As introduced in section 3.4 and further discussed in Appendix B the stability of the solution can be inferred from the term multiplying $\ddot{\phi}$ in eq. (3.58). For this theory, the factor we are concerned with is

$$Z^{tt}_{\rm MG} = -3 - \alpha\left(2T^{00} - 6\phi'' - 12\frac{\phi'}{r}\right) - \alpha^2\left(\frac{6(\phi')^2}{r^2} + \frac{12\phi''\phi'}{r}\right)\,. \tag{3.65}$$

The requirement for the stability of the static solution (3.63) against high-frequency modes can be shown to be equivalent to require $Z^{tt}_{\rm MG} < 0$ at all points in spacetime, which can be used to put constraints on the theory's free parameter α. Analyzing the $Z^{tt}_{\rm MG}$ factor leads to the realization that it is possible to have $Z^{tt}_{\rm MG} > 0$ inside the source for negative values of α. The novel coupling of ϕ to the energy-momentum tensor plays a key role for this to happen. Note that in the background of a static solution, $Z^{tr}_{\rm MG} = 0$, so the condition for the Cauchy breakdown $Z^{tt}_{\rm MG} = 0$ is also the condition for the solution to be marginally stable. In general $Z^{tt}_{\rm MG} < 0$ for all positive α and the spatial components of the effective metric $Z^{rr}_{\rm MG}$ and $Z^{\theta\theta}_{\rm MG}$ are positive for all values of α (as long as we neglect pressure), so our only concern is that $Z^{tt}_{\rm MG}$ becomes positive when α is negative. Setting $R_0 = 1$ in eq. (3.7) and using the relation $\phi' = \lambda_1 r$, the term $Z^{tt}_{\rm MG}$ can be written in terms of a single parameter $\kappa \equiv \alpha\rho$. One finds that

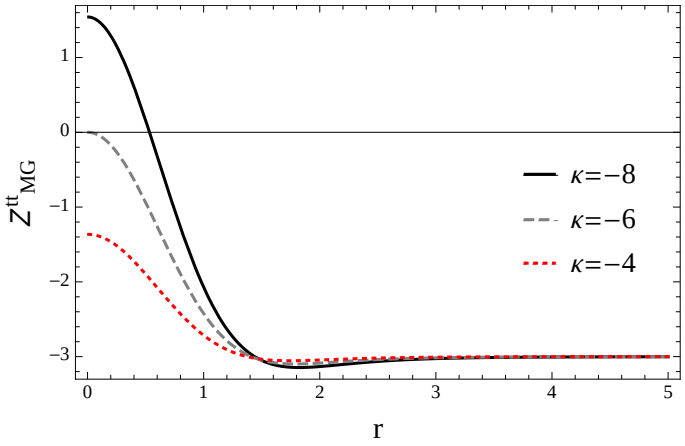

Figure 3.15: The factor $Z^{tt}_{\rm MG}$ for various values of $\kappa \equiv \alpha\rho$. The solution is unstable for $\kappa < -6$.

$Z^{tt}_{\rm MG} > 0$ inside the source for $\kappa < -6$. Therefore, as long as $\alpha > -6/\rho$, the solution is stable, as shown in figure 3.15.

For physically realistic values ($\rho \sim 10^{26}$ for a sun-like source), the window of stability $-6/\rho < \alpha < 0$ is quite small so we conclude that the only valid static screening solution in the massive gravity decoupling limit is the one with cosmological asymptotics. As shown in [191], this solution is stable against linear perturbations. In addition, fluctuations remarkably propagate with sub-luminal speeds.

We have thus seen how the extra free parameter α and the new coupling $\nabla_\mu\nabla_\nu\phi T^{\mu\nu}$ give some qualitative differences to the static spherically symmetric solutions in massive gravity as compared to DGP. But at this point, the study of dynamical solutions is quite similar to the DGP case, except we only have one branch of static solution to analyze. We can use the same numerical method to solve the time-dependent equation (3.58) and perform the same tests as we did in the DGP model.

3.7.2 Dynamical stability analysis

An analysis of the non-linear stability of the static screening solution (3.63) revealed what we expected from our detailed study of the DGP model: the static solution is reached as the endpoint of the evolution for the gaussian wavepacket considered in Section (3.3.2), as long as the fluctuation is small enough compared to the background solution; the static solution is also reached considering the quadratic vacuum initial conditions described in Section (3.3.2), as long as R_V/R_0 is sufficiently small.

However, for large perturbations we found some different qualitative features which are due

mainly to two reasons: the additional coupling to matter $\nabla_\mu\nabla_\nu\phi T^{\mu\nu}$; and the big hierarchy between $Z^{\theta\theta}$ and Z^{tt} of the effective metric, $Z^{tt} \gg Z^{\theta\theta}$ [125]. The components of the effective metric can be found in eq. (3.57).

Due to the extra coupling to matter the theory is less prone to suffer Cauchy breakdown near the source than in DGP. This can be traced back to the fact that inside the source, fluctuations are highly suppressed due to this term [191]. In fact, when pressure is neglected, $Z^{rr}_{\rm MG}$, $Z^{\theta\theta}_{\rm MG}$ will always change sign before Cauchy breakdown occurs due to the big hierarchy between the different components of the metric $Z^{\mu\nu}$. The conditions for this to occur are qualitatively similar to the ones we found in Section 3.4. For very large fluctuations, we were not always able to evolve past these points, most likely due to the excitation of high-frequency unstable modes. However, we expect that the formation of these unstable regions causes an enhancement of gradients of the field, making Cauchy breakdown inevitable. Cauchy breakdown was more easily observed outside the source where we checked that the condition (3.41) was satisfied, showing once again that Cauchy breakdown is a coordinate independent phenomenon.

In conclusion, even though the extra coupling to matter renders the theory less prone to Cauchy breakdown inside the source, problems can arise for sufficiently large perturbations of the screening solutions (3.63). On the other hand, for the class of well-behaved initial conditions the screening solution is stable and behaves as a coherent object with radius R_V, as we discuss below.

Quasinormal modes and tails

As discussed before in the DGP case, when perturbed, the static solutions (3.63) vibrate and eventually relax to an equilibrium state again. Performing the same analysis as in Section 3.3.3, we find that the waveform consists of the expected three stages, a prompt response at very early times, quasinormal modes at intermediate times and a power-law tail at very late times.

We saw in the previous section that the new coupling $\nabla_\mu\nabla_\nu\phi T^{\mu\nu}$ is important for the stability of the solutions. If this coupling is absent the behavior of perturbations on top of the static solution (3.63), for $\alpha < 0$ and $\alpha > 0$ is very similar to the DGP case (see figure 3.5). The introduction of this new coupling makes the solution for $\alpha < 0$ unstable, which can be clearly seen in a time-domain analysis of the linear equation around this background. On the other hand perturbations on top of the asymptotically growing solution with $\alpha > 0$, are stable and have a clear quasinormal ringdown similar to the one shown in the top panel of figure 3.5.

A frequency domain analysis also shows that for $R_V \gg R_0$ the quasinormal frequencies

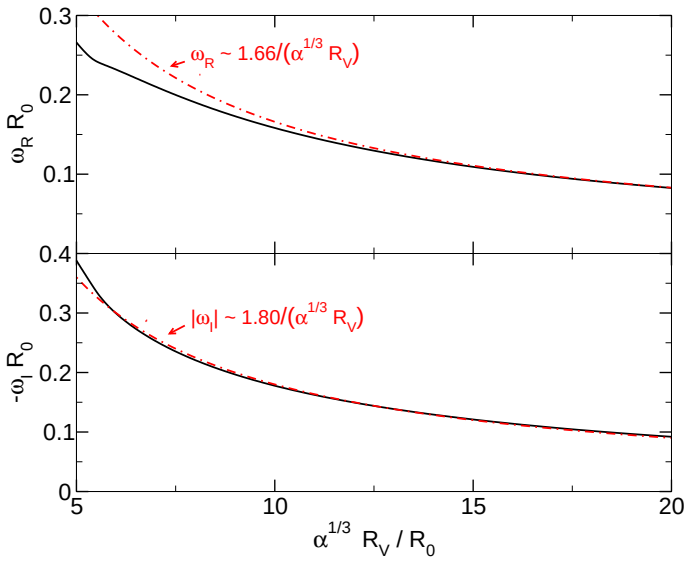

Figure 3.16: Fundamental quasinormal modes of the scalar field in the decoupling limit of massive gravity. The full lines correspond to the numerical results, whereas the dashed lines show the analytic approximation at low frequencies. The top and bottom panels show the real part, $\omega_R R_0$, and the imaginary part, $\omega_I R_0$, of the mode as a function of $\alpha^{1/3} R_V / R_0$.

follow the same trend as in DGP and are given by

$$\omega_R \sim \frac{1}{\alpha^{1/3} R_V}\,, \tag{3.66}$$

$$\omega_I \sim -\frac{1}{\alpha^{1/3} R_V}\,. \tag{3.67}$$

This is shown in figure (3.16), where we plot the fundamental quasinormal modes. One can understand this scaling from the fact that the coupling constant α can be reabsorbed into Λ_3 and so the effective Vainshtein radius is given by $\tilde{R}_V \equiv \alpha^{1/3} R_V$.

In the background of (3.63), we can compute a wave equation of the form (3.26), where the effective potential has the large distance asymptotic behavior

$$V \sim \frac{l(l+1)}{r_*^2} + \frac{K}{r_*^8}\,, \tag{3.68}$$

where K is once more a constant that depends on l and $M(r \to \infty)$. This behaviour is independent of α and does not depend on the new coupling $\nabla_\mu \nabla_\nu \phi T^{\mu\nu}$ or the specific form of $T^{\mu\nu}$. Thus, the above analysis suggests that scalar perturbations of the static solution generically decay as $\psi(r_*, t) \sim t^{-2l-8}$ at late times, just as in DGP.

3.7.3 Collapsing and Exploding Sources

Compared to DGP, the extra coupling to matter in massive gravity makes the scalar evolution less prone to Cauchy breakdown, although it can not be avoided for sufficiently large fluctuations.

Does this hold for the dynamical sources that we considered in Section 3.5? In general yes, although some additional subtle issues are worth pointing out.

For the case of the collapsing source (3.43), when pressure becomes important, unstable regions near the source are inevitable, as was pointed out in [177]. Instabilities as well as sound horizons form during collapse as the pressure becomes significant $p \sim \rho$, as can be inferred by looking at eq. (3.57). The spatial components, and in particular $Z^{\theta\theta}$, will change sign at some point in space, in a finite time. Although this behavior was also found in DGP, due to the extra coupling to matter, this effect is enhanced here. However, this is in general followed by Cauchy breakdown which makes it impossible to follow the development of the instability. This is not only dependent on the density of the source but also on the decay time–scale, with a relation similar to the one found in DGP. For large time–scales (or very small densities) the field will evolve without instability or breakdown to a different solution with the same asymptotics (not to the vacuum solution, due to the extra coupling to matter). But for realistic source densities, Cauchy breakdown seems to be inevitable.

Surprisingly, the exploding source (3.44) seems to avoid Cauchy breakdown for moderately high densities $\rho \sim 10^6$ (and $\alpha \sim 1$, recall that α can be reabsorbed). Once more, this is mainly due to the hierarchy $Z^{tt} \gg Z^{ii}$ in this background. However, for sufficiently large α and ρ, regions where the eigenvalues of Z^{rr} and $Z^{\theta\theta}$ change sign can form. In these regions, the field fluctuations propagate at extreme subluminal velocities, thus leaving time for instabilities to grow. Similar to the above cases, these instabilities can eventually cause Cauchy breakdown.

3.7.4 Asymmetric screening solutions

Let us close our discussion on massive gravity by following Section (3.6) to compute asymmetric screening solutions for the source (3.51). These turn out to be very similar to the ones found in the DGP model. The asymptotic form of the scalar multipole components are the same as in DGP, namely $\delta\phi \sim A_1 r^l$ at the origin and $\delta\phi \sim r^{l+1}$ at infinity. Some solutions are shown in figure 3.17. For $R_0 \ll r \ll R_V$ the field decays as $r^{-l/4}$ indicating that in the Vainshtein regime higher multipoles have a stronger suppression. The suppression of their contribution to the fifth force compared to the *multipoles* of the Newtonian gravitational force $F_g \sim r^{-(2+l)}$ is given by

$$\left|\frac{F_{l>0}}{F_g}\right| \sim \left(\frac{r}{R_V}\right)^{1+3l/4} \quad \text{if } R_0 \ll r \ll R_V \,. \tag{3.69}$$

This shows that in this case, in the Vainshtein regime solutions are generically more suppressed in massive gravity than in DGP.

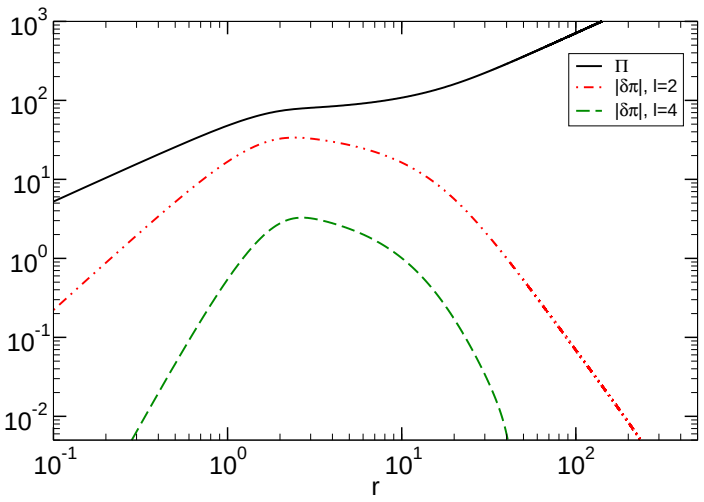

Figure 3.17: Hairy solutions for different multipoles l, here shown for $R_V = 100$, $z_0 = 0.9$, $\alpha = 1/3$, and compared to $\Phi \equiv \phi'$. For $R_0 \ll r \ll R_V$ the field decays as $\rho^{1/3} z_0^l r^{-l/4}$ while for $r \gg R_V$ they decay as $\rho z_0^l r^{-(l+1)}$

3.8 Conclusions

The theoretical challenge posed by explaining the observed accelerated expansion of the Universe has put theoretical physics at a crossroads. One can postulate dark energy as the cause, possibly in the form of a cosmological constant, and be content with an anthropic explanation for the accelerated expansion. On the other hand, one can question the validity of GR on large distance scales, and be left with the need to explain why local departures have not been observed. In this chapter, we have explored the latter possibility, studying the linear and non-linear stability of the screening solutions that restore the predictions of GR on short distance scales.

The two theories we have studied, the DGP braneworld scenario and dRGT massive gravity, are examples where non-linear derivative interactions give rise to modifications of gravity only in the infrared through the Vainshtein screening mechanism. We have focused on the decoupling limit of these theories, in which a non-gravitating scalar degree of freedom is introduced that couples to the matter sector. Using analytic and numerical methods, we have taken the first steps towards establishing the fully non-linear dynamical stability of the Vainshtein screening solutions in spherical symmetry. We have also derived some properties of screening solutions beyond spherical symmetry. Our main results are as follows:

- Using numerical simulations we have shown for the first time that in the decoupling limit of both DGP and dRGT massive gravity, the Vainshtein screening solutions are dynamically accessed from a wide variety of initial conditions beyond the linear regime.

- We have shown that the screening solutions behave as a coherent object much like a star or black hole under linear perturbations: a prompt response due to the primary scattering is followed by a universal series of damped oscillations known as quasinormal modes, which is then followed by a universal power-law decay. This analysis also shows that spherical sources can only support a monopole configuration of the scalar π; any multipolar "hair" on spherically symmetric screening solutions is radiated away.

- However, for sufficiently large perturbations, regions of spacetime form in which there is no longer a well-defined Cauchy problem, a phenomenon which (following previous nomenclature) we term Cauchy breakdown. We have shown that in general this is not a coordinate singularity, but a real physical problem. In the absence of a new physical principle for what occurs in such regions, the future evolution is undetermined. This phenomenon is a general feature of theories with derivative self-interactions.

- For sources which undergo collapse into a relativistic object or explode, we have shown that Cauchy breakdown generically occurs when there is a large hierarchy between the radius of the source and the Vainshtein radius. This is the case for realistic astrophysical objects, and hence there is the danger that Cauchy breakdown will occur in a complete description of violent astrophysical phenomena such as supernovae or the formation of neutron stars and black holes.

- Finally, by considering non-spherically symmetric sources, we have shown that for both DGP and dRGT massive gravity the tidal components of the scalar fifth force are subdominant to the gravitational tidal field, and that tidal forces are screened more effectively than the monopole.

Our results represent a first step towards establishing the nonlinear dynamical stability of infrared modifications of gravity. However, Cauchy breakdown is likely to be an important obstruction to determining the stability of a variety of cases of potential physical interest. What might be necessary to determine evolution past these points? In this work, we have neglected quantum corrections, which by a naive analysis become relevant whenever derivatives of the π field become sufficiently large compared to the scale Λ_3.

However, the Vainshtein mechanism itself changes the scale at which fluctuations become strongly coupled. On top of a background configuration, the strong coupling scale gets "redressed" by the effective metric and is given by $\tilde{\Lambda_3} \equiv \Lambda_3 Z^{1/2}$ [41], where Z schematically represents the relative strength of the eigenvalues of a slowly varying effective background $Z^{\mu\nu}$. For DGP and dRGT massive gravity, Z can symbolically be written as $Z \sim 1 + \partial^2\pi_0/\Lambda_3^3$. In the non-linear regime $r \ll R_V$, small fluctuations around the static screening solutions see

an effective metric with $Z \gg 1$, implying that $\tilde{\Lambda_3} \gg \Lambda_3$. Thus, for the static backgrounds that we considered, quantum corrections are suppressed. However, for big fluctuations around the static solutions the condition $Z \gg 1$ does not hold in general, leaving the possibility for Cauchy breakdown to occur, in which case $\tilde{\Lambda_3} \to 0$. This means that at this point fluctuations become infinitely strongly coupled [207], signaling that the classical theory can no longer be trusted. The scale Λ_3 (and the redressed scale $\tilde{\Lambda_3}$) is the strong coupling scale of the theory, but not necessarily the theory's cutoff. Hitting the scale Λ_3 or $\tilde{\Lambda_3}$ does not automatically imply a breakdown of the physical theory, but rather a breakdown of perturbativity. This means that quantum loops should be taken into account, but it does not necessarily mean that new physics is required. On the other hand, understanding how to evolve past Cauchy breakdown may require an understanding of how infrared modifications of gravity can be UV completed [3].

Outside the regime of Cauchy breakdown, it is possible to study a variety of situations of physical interest. In particular, the formation of cosmological large scale structure, and perhaps some solutions in the strong field regime. In future work, we will tackle these problems, extending our analysis beyond the decoupling limit, and considering evolution that includes the internal dynamics of realistic sources. We hope that this work will produce new predictions for observables, aiding in the search for the cause of the observed accelerated expansion of the Universe.

[3]It is curious to note that our results clearly show that the screening solutions behave as coherent objects extended up to R_V, which could be closely related to some proposals for the UV completion of these theories [213, 214, 215].

Chapter 4

Gravitational waves in massive bigravity

In this chapter we study a primordial stochastic gravitational wave background in massive bigravity. The phenomenology can differ from standard General Relativity due to non-trivial mixing between the two linearized tensor fluctuations in the theory, only one of which couples to matter. We study perturbations about two classes of cosmological solutions in bigravity, computing the tensor contribution to the temperature anisotropies in the Cosmic Microwave Background radiation and the present stochastic gravitational wave background. The result is strongly dependent on the choice of cosmological background and initial conditions. One class of background solution remains observationally indistinguishable from standard General Relativity for a wide variety of initial conditions, while the other generically displays tremendous growth in the amplitude of large-wavelength gravitational waves. We analyze the initial conditions for tensor modes expected in an inflationary cosmology. We find that for the branch with growing perturbations, inflation generically yields initial conditions that sufficiently suppress the growth. When evolved, this gives rise to a stochastic background observationally indistinguishable from standard General Relativity.

4.1 Introduction

Massive bigravity, introduced in section 2.4, is a theoretically consistent modification of general relativity with an additional dynamical rank two tensor. As for any modified gravity theory, much of the interest in massive bigravity is motivated by the puzzle of cosmic acceleration and dark energy. Viable homogeneous and isotropic cosmological solutions exist in bigravity which can describe our universe, including acceleration, without a cosmological constant [158, 159, 161, 162]. These are usually referred to as *self-accelerating* background solutions. Beyond the background level, investigations have recently been underway to analyze perturbations in bigravity[165, 166, 160, 158, 163, 216, 217]. These analyses reveal that there is a particular class of stable solutions, but all others are plagued by an exponential instability in the scalar sector in the early universe [167].

The transverse traceless fluctuations of each of the two dynamical metrics in bigravity interact, altering the propagation of gravitational waves as compared to GR. The most general set of linearized equations of motion for the visible sector tensor modes h_g and the dark sector tensor modes h_f is given by

$$\mathbf{D}^2 \cdot \mathbf{h} + \mathbf{m}^2(\mathbf{x}, \tau) \cdot \mathbf{h} = 0 \tag{4.1}$$

where

$$\mathbf{h} = \begin{pmatrix} h_f \\ h_g \end{pmatrix}, \quad \mathbf{D}^2 \equiv \begin{pmatrix} \nabla_f^2 & 0 \\ 0 & \nabla_g^2 \end{pmatrix}, \quad \mathbf{m}^2(\mathbf{x}, \tau) \equiv \begin{pmatrix} m_f^2(\mathbf{x}, \tau) & m_{fg}^2(\mathbf{x}, \tau) \\ m_{gf}^2(\mathbf{x}, \tau) & m_g^2(\mathbf{x}, \tau) \end{pmatrix}, \tag{4.2}$$

where ∇_f^2 is the covariant derivative defined with respect to $f_{\mu\nu}$ and ∇_g^2 is the covariant derivative defined with respect to $g_{\mu\nu}$. There is no off-diagonal term in the differential operator $\mathbf{D}^2$ due to the absence of consistent derivative couplings between the two metrics. The mass matrix $\mathbf{m}^2$ is in general not diagonal or symmetric. The differential operator $\mathbf{D}^2$ and mass matrix $\mathbf{m}^2$ are generally not simultaneously diagonalizable, leading to mixing between the visible and dark sector tensors. Because the visible and dark sector tensor fluctuations evolve in a different background and possess a different mass, their propagation speeds can in general be different and spacetime dependent. The modifications in the propagation speed of gravitational waves and mixing between the visible and dark sector tensors generically present in bigravity can have important implications for both astrophysical [218, 219, 220] and primordial [221, 222, 66, 223, 224] gravitational waves.

In this chapter, we consider a stochastic background of primordial gravitational waves in massive bigravity, studying the impact on primordial and present day gravitational wave observables by computing the tensor contribution to the temperature anisotropies of the Cosmic Microwave Background radiation (CMB) and the power spectrum of the present day

stochastic gravitational wave background. We examine whether these observables can be used to distinguish between bigravity and GR, serving as a possible test for gravity on large cosmological scales. We consider two classes of cosmological background solutions for the dark sector metric $f_{\mu\nu}$. For one class, the tensor perturbations match those of GR very closely. However, for the second class of background solution, a power-law instability in the tensor sector leads to a strong growth in the amplitude of gravitational waves at late times.

In this work we discovered that the power-law instability in the dark gravitational wave sector of the bouncing branch leads to physical gravitational waves that grow in amplitude with time, consistent with the findings in refs. [216, 217]. Since gravitational waves are supposed to decay, this growth would be enough to rule out this branch of models. However, the amount of growth is extremely sensitive to the initial conditions, so without a proper analysis of initial conditions, the fate of the theory remains unknown. Our analysis addresses this outstanding question regarding the viability of massive bigravity by explicitly exploring the initial conditions of bigravity in the context of inflation. We show that inflationary initial conditions do not excite the growing mode significantly. In fact, our analysis proves that in the context of inflation, the growing mode is so severely suppressed that there will be no observable deviation from GR in the CMB or stochastic gravitational wave background. These results motivate a more in depth consideration of inflationary model building in massive bigravity.

The chapter is organized as follows: following our description of bigravity cosmology from section 2.4 we formalize the analysis of the gravitational waves in each background configuration in section 4.2. We then compute gravitational wave observables: section 4.3 gives the CMB Tensor Power Spectrum in bigravity, and section 4.4 gives the present day stochastic gravitational wave background. Section 4.5 explores the initial conditions as predicted by inflation, followed by a discussion in 4.6.

4.2 Bigravity Tensor Perturbations

Here we give the equations of motion for the transverse traceless tensor modes $h^{\mathrm{TT}}_{g,ij}$ and $h^{\mathrm{TT}}_{f,ij}$ corresponding to the metrics $g_{\mu\nu}$ and $f_{\mu\nu}$ respectively. To compute observables, we will be most interested in the perturbations corresponding to the physical metric, $h^{\mathrm{TT}}_{g,ij}$, since these are the ones that couple to matter. The tensor equations of motion were first analyzed in [165]. A thorough analysis of scalar, vector, and tensor perturbations was performed in [163]. In addition, further analysis of tensor perturbations appeared in Refs. [216] and [217]. The

tensor perturbation equations of motion in momentum space are

$$\ddot{h}_g + 2\mathcal{H}\dot{h}_g + k^2 h_g + a^2\lambda(h_g - h_f) = 0 \tag{4.3}$$

$$\ddot{h}_f + \left[2\left(\mathcal{H} + \frac{\dot{r}}{r}\right) - \frac{\dot{c}}{c}\right]\dot{h}_f + c^2 k^2 h_f - \frac{a^2\lambda c}{r^2}(h_g - h_f) = 0 \tag{4.4}$$

where superscripts and subscript indices have been dropped for simplicity. In addition, the time-dependent function λ is defined as

$$\lambda = \beta_3 c r^3 + \beta_2(c+1)r^2 + \beta_1 r \tag{4.5}$$

which simplifies to $\lambda = \beta_1 r$ in either branch under consideration. These equations are satisfied separately for each polarization; the polarizations do not mix.

Initial conditions for h_g and h_f at some initial time τ_i are required to obtain solutions. In the absence of a theory of initial conditions, we should consider general initial data $h_{(g,f)}(\tau_i)$ and $\dot{h}_{(g,f)}(\tau_i)$. However, for standard inflationary cosmology in GR, tensor modes freeze in once their physical wavelength becomes comparable to the primordial horizon size, motivating $\dot{h}_g(\tau_i) = 0$. In section 4.5, we compute the initial conditions expected for inflationary cosmology in the context of bigravity, finding that $\dot{h}_{(g,f)}(\tau_i) = 0$ is an appropriate choice. Note that this assumption differs from Cusin et. al., who consider initial data with $\dot{h}_f(\tau_i) \neq 0$. This was motivated by the presence of a growing mode for h_f, which if excited, would dominate the evolution. Our choice of initial data initially sets this growing mode to zero, and as shown below, this leads to a different growth history for h_f at late times.

Expanding Branch

In the expanding branch, the factor $\left[2\left(\mathcal{H} + \frac{\dot{r}}{r}\right) - \frac{\dot{c}}{c}\right]$ is always positive, and causes significant damping of the dark sector tensor perturbation h_f. In addition, the factor $a^2\lambda$ is small in this branch, $a^2\lambda < 0.3$ for all time. Therefore, unless the initial amplitude of h_f is very large compared with h_g, the mixing term $a^2\lambda(h_g - h_f)$ does not significantly alter the behaviour of the physical tensor perturbation h_g. For equal initial amplitude, our numerical solutions for the tensor modes in this branch match closely with those of pure GR. This was confirmed for modes ranging from $k = 0.1H_0$ to $k = 10^5 H_0$. Refer to figure 4.1 for the numerical solutions of this branch as compared to the standard GR gravitational waves with $\tau_i = 10^{-6}H_0^{-1}$. [1]

[1] Although this is choice for τ_i, corresponding to reheat temperature $T_i = T_{\rm eq}[a_{\rm eq}/a(\tau_i)] = T_0(a_0/a_{\rm eq})^2[a_{\rm eq}/a(\tau_i)] = 0.07$ GeV, is not entirely plausible, it is still deep within the radiation era and is a practical choice for our numerical analysis. We will be able to extrapolate to earlier τ_i when necessary using scaling properties of the solutions defined in (4.7).

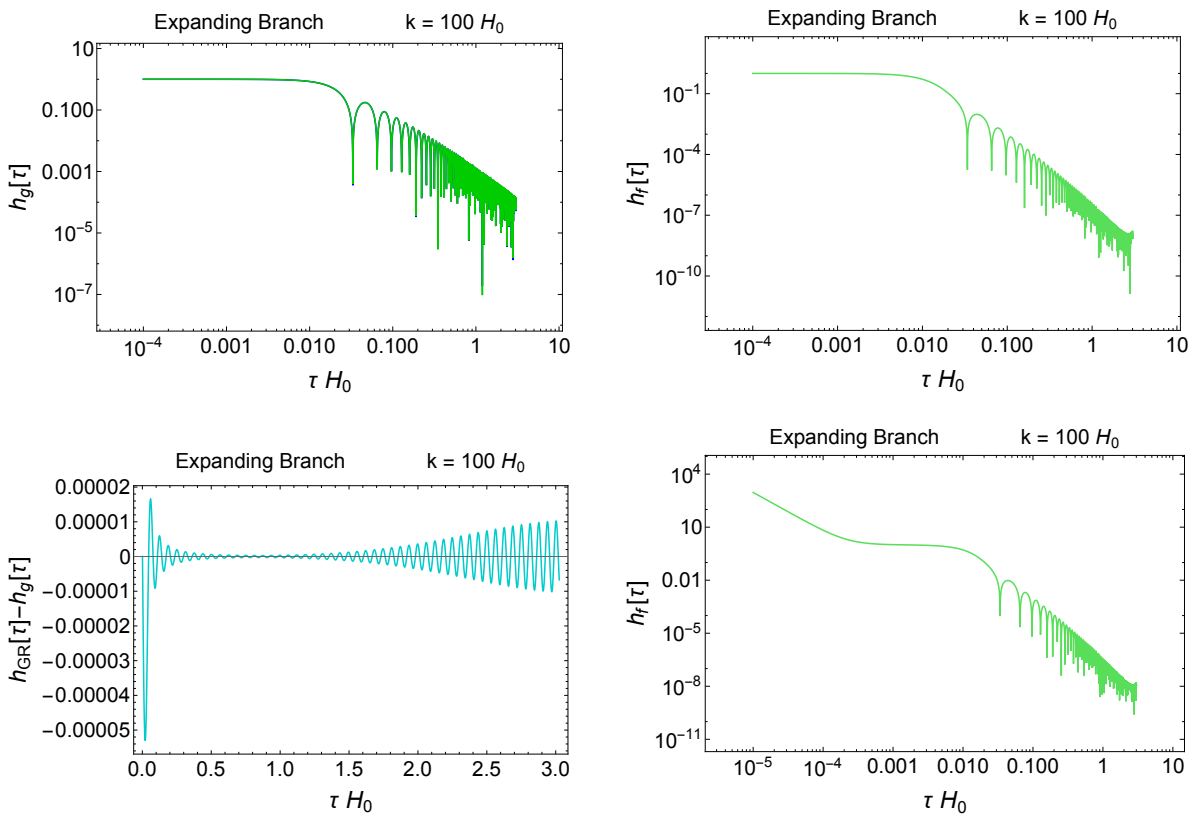

Figure 4.1: Top left: the solution for $h_g(\tau)$ at $k = 100H_0$ for the expanding branch (green) versus GR (blue). The two solutions are essentially indistinguishable. Top right: the solution for $h_f(\tau)$ at $k = 100H_0$ for the expanding branch. Bottom left: The difference between the expanding branch solution and the standard GR solution, shown to be less than 5×10^{-5} for $k = 100H_0$. Bottom right: the solution for $h_f(\tau)$ at $k = 100H_0$ for the expanding branch, with $h_f(\tau_i)/h_g(\tau_i) = 10^6$. Even in this case of $h_f(\tau_i) \gg h_g(\tau_i)$, the decay of h_f is so fast that it does not cause any alteration to the physical tensor perturbations.

If the relative amplitude of the dark sector tensor mode was decreased, $h_f(\tau_i) < h_g(\tau_i)$, this would only drive the physical tensor modes closer to those of GR. But if h_f had a much higher initial amplitude relative to h_g, the mixing term $a^2\lambda h_f$ in (4.4) can become dominant for some time. However, even if we set $h_f(\tau_i) \gg h_g(\tau_i)$, h_f decays so dramatically that there is very little effect on h_g. The bottom right plot in figure 4.1 shows how the dark sector perturbation decays very quickly, even when starting with a much larger amplitude.

Bouncing Branch

In the bouncing branch, the oscillations of h_f are anti-damped for $\tau < \tau_b$ and damped for $\tau > \tau_b$ where τ_b is the bounce time corresponding to when $c(\tau) = 0$. This happens at relatively late times, around $z_b \sim 0.6$. Until the bounce occurs, while $c < 0$, h_f experiences enormous growth, then starts to decay after τ_b. Through the mixing term in (4.3), this growth in h_f

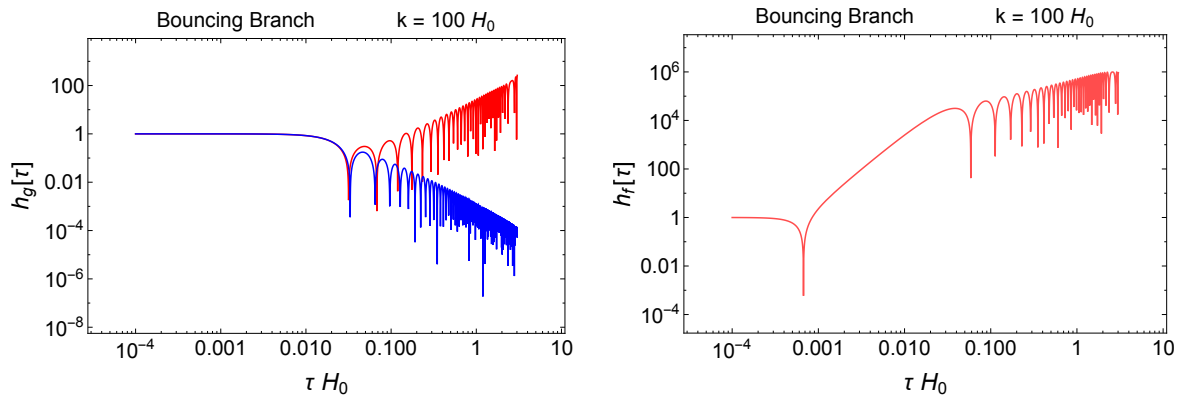

Figure 4.2: Left: the solution for $h_g(\tau)$ at $k = 100H_0$ for the bouncing branch (red) versus GR (blue). Right: the solution for $h_f(\tau)$ at $k = 100H_0$ for the bouncing branch.

translates into growth in the physical mode h_g, leading to oscillations that grow at late times. In figure 4.2, we show the evolution of the physical and dark sector tensors for equal amplitude initial conditions with $\tau_i = 10^{-6}H_0^{-1}$; the large deviation from GR at late times is clear.

The growth in $h_g(\tau)$ at late times is k-dependent. Empirically, we find a falloff proportional to $1/k^2$. This amplification is also dependent on the initial time τ_i, which physically relates to the reheating temperature $T_{\rm reheat} = T_i$. This can be understood by examining the solution for h_f within the radiation dominated era. On super-horizon scales there is an exact solution [163] given by

$$h_f = c_1 + c_2\tau^3 \tag{4.6}$$

so there is both a constant mode and a growing mode. Our choice of initial condition $\dot{h}_f = 0$ selects the constant mode (ie. $c_2 = 0$), so naively one might think that h_f should not grow at all in the radiation dominated era, regardless of the initial time τ_i. However, the above solution (4.6) is only an approximate solution on super-horizon scales, not valid for $k \neq 0$. For $k \neq 0$ we expect to depart from the the constant mode solution on a timescale of $\tau \sim 1/ck$, at which point the growing mode will completely dominate. The earlier the initial time is set, the more time h_f has to grow, ultimately driving more growth in h_g. We find that the growth is inversely proportional to the initial time: $h_g(\tau_0) \propto 1/\tau_i \propto T_i$, valid for all initial times in the radiation dominated era.

In addition to this, the growth is also proportional to the initial condition for h_f. However, for a small enough value of $h_f(\tau_i)$, the solution for $h_g(\tau)$ does not scale. For our choice of initial time $\tau_i = 10^{-6}H_0^{-1}$ (corresponding to a reheat temperature of $T_i = 0.07$ GeV) we find that for $h_f(\tau_i)/h_g(\tau_i) < 10^{-9}$, the solution for h_g is indistinguishable from its solution with $h_f(\tau_i) = 0$ which agrees very closely with the pure GR solution. To extrapolate this

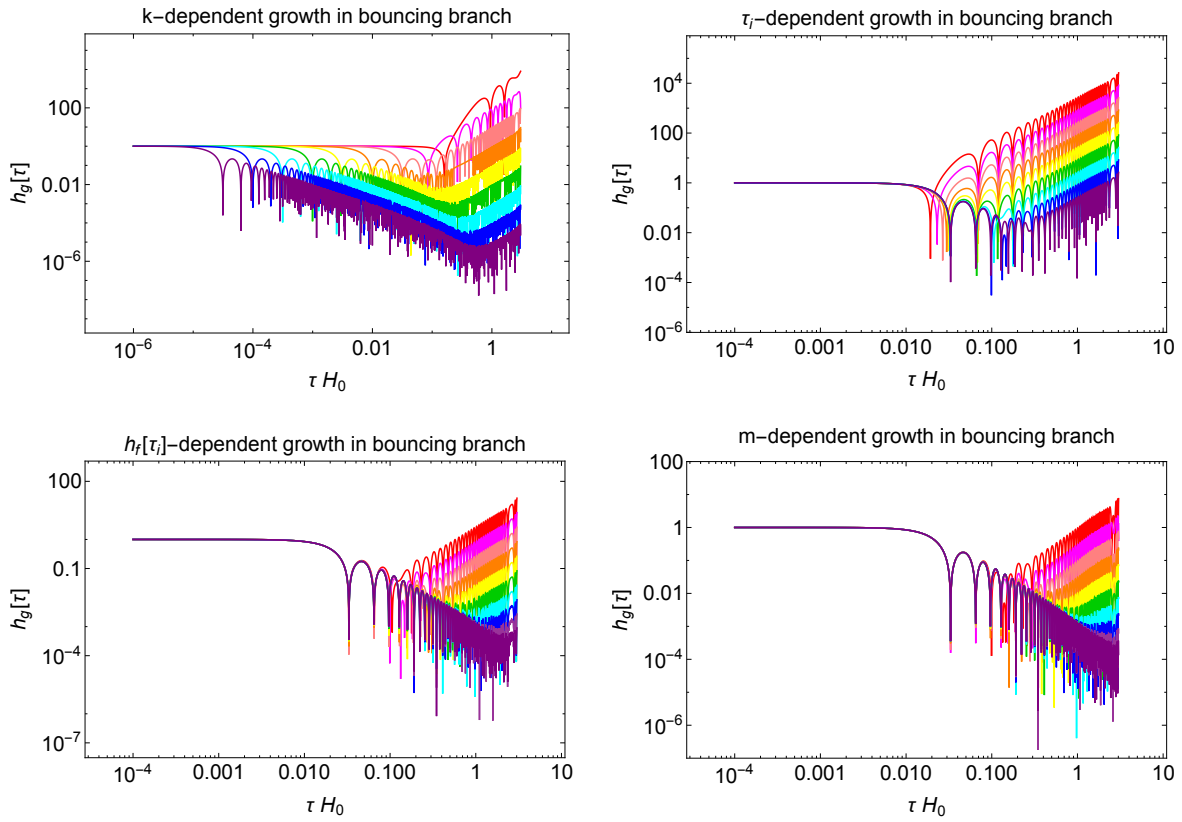

Figure 4.3: The growth in the physical sector is dependent on k, τ_i, $h_f(\tau_i)$ and m (4.7). We vary one parameter individually per plot to show how the solutions change, fixing all but one of $\tau_i = 10^{-6}H_0^{-1}$ ($T_i \sim 0.07$ GeV), $k = 100H_0$, $m = H_0$, and $h_{(g,f)}(\tau_i) = 1$. Top left: The gravitational waves in the bouncing branch grow like $1/k^2$. Here we plot $k = 10^{p/2}H_0$ with p ranging from 2 to 10 from top to bottom. Top right: The tensor modes grow like $1/\tau_i$. Here we plot $\tau_i = 10^{-8+p/2}H_0^{-1}$ with p ranging from 0 to 8 from top to bottom. Bottom left: The gravitational waves grow proportional to $h_f(\tau_i)$. Here we plot $h_f(\tau_i) = 10^{-p/2}$ with p ranging from 2 to 11 from top to bottom. Bottom right: The growth in the bouncing branch grow proportional to m (requires re-inserting Λ into the theory). Here we plot $m = 10^{-p/2}H_0$ with p ranging from 2 to 11 from top to bottom.

result to a more reasonable reheat temperature, say $T_i = 10^{10}$ GeV, we must consider that h_f grows proportionally to T_i. We conclude that for $T_i = 10^{10}$ GeV, we need a suppression of $h_f(\tau_i)/h_g(\tau_i) < 10^{-20}$ to obtain solutions that agree with those of GR. At this threshold, scaling down $h_f(\tau_i)$ further will not result in any significant change. It is evident that to control the large growth in the bouncing branch, seeking gravitational waves that do not substantially deviate from those of GR, we require detailed knowledge about the mechanism by which they were produced. This will be explored further in section 4.5.

From a purely phenomenological standpoint, one might also be interested in varying the graviton mass. This reduces the strength of the mixing term, which has the form $\propto m^2\beta_1^* a^2 r(h_g - h_f)/H_0^2$ (reinserting factors of m and H_0 from (2.41)). However, lowering m means weakening the influence of the bigravity interaction term in (3.1) to the point at

which it can no longer yield a viable background cosmology, so a cosmological constant must be reintroduced. This undercuts the strongest theoretical motivation for bigravity, but mild variations could be of interest in constraining the parameters of the theory. We find that $h_g(\tau_0)/h_g(\tau_i) \propto m$. Just like for $h_f(\tau_i)$, there is also a critical value of m for which this scaling relation no longer holds. Putting this all together, we find the following general relationship for the growth:

$$\frac{h_g(\tau_0)}{h_g(\tau_i)} \propto \frac{m h_f(\tau_i)}{\tau_i k^2} \propto \frac{m h_f(\tau_i)\mathcal{H}(\tau_i)}{k^2} \propto \frac{m h_f(\tau_i) T_i}{k^2}, \qquad \text{for } \frac{h_f(\tau_i)}{h_g(\tau_i)},\ m/H_0 > A_{\text{crit}} \tag{4.7}$$

where A_{crit} is the critical value at which the solution no longer scales ($A_{\text{crit}} = 10^{-20}$ for $T_i = 10^{10}$ GeV). Note that this scaling is different than that found in Cusin et. al., which is due to the difference in choices for initial data. These growth dependencies are displayed in figure 4.3 which shows the solutions for various k, τ_i, $h_f(\tau_i)$, and m. Notice that the variations in growth of h_g are similar for $h_f(\tau_i)$ and m, which comes from the fact that both of these effects act to alter the coupling term in (4.3). Although the effect on growth in the physical sector appears the same, altering m introduces no change in growth in the dark sector.

4.3 CMB Tensor Power Spectrum

In this section we use the solutions of (4.3) and (4.4) to compute the predicted bigravity CMB temperature tensor power spectrum. We begin by using the physical tensor mode $h_g(k,\tau)$ to compute the lth photon moment due to tensor perturbations (assuming instantaneous recombination) using

$$\Theta_l^T = -\frac{1}{2}\int_{\tau_*}^{\tau_0} d\tau\ \dot{h}_g(k,\tau) j_l[k(\tau_0 - \tau)]] \tag{4.8}$$

where τ_0 is the conformal time today, τ_* is the conformal time at the time of last scattering, and j_l is the spherical Bessel function. The tensor contribution to the temperature anisotropies is entirely due to the Integrated Sachs Wolfe effect. The evolution of the visible sector tensors can be substantially different than GR, leading to potentially visible differences in the spectrum of temperature anisotropies. For example, in the bouncing branch, from figure 4.3 it can be seen that there is additional time dependence on super horizon scales, and late-time growth of tensor modes; both of these effects will alter the temperature anisotropies.

Once the photon moments are found, the angular power spectrum can be found from:

$$C_l^T = \frac{(l-1)(l+1)(l+2)}{\pi}\int_0^\infty dk\ \frac{1}{k}\left|\frac{\Theta^T_{(l-2)}}{(2l-1)(2l+1)} + 2\frac{\Theta^T_l}{(2l-1)(2l+3)} + \frac{\Theta^T_{(l+2)}}{(2l+1)(2l+3)}\right|^2$$

We solve for the photon moments and angular power spectra numerically.

For the expanding branch, the C_l^T's are approximately the same as in GR, which is expected given the agreement of $h_g(\tau)$ with $h_{\rm GR}(\tau)$. However, in the bouncing branch, there is a drastic difference in the power spectrum.

In (4.8) the integral runs from the time of last scattering to today, so the growth in $h_g(\tau)$ in the bouncing branch at late times causes a large increase in Θ_l^T and therefore in C_l^T. We find that if we set $h_g(\tau_i) = h_f(\tau_i) = 1$ at $\tau_i = 10^{-6}H_0^{-1}$ and $m = H_0$, then the value of C_2^T for the bouncing branch is 10^{11} times larger than for pure GR: $C_{2\,\rm BB}^T \sim 10^{11} C_{2\,\rm GR}^T$, demonstrating that the growth at late times dominates the signal. However, our initial time $\tau_i = 10^{-6}H_0^{-1}$ corresponds to a reheat temperature of only $T_i = 0.07$ GeV. Extrapolating to a more reasonable reheat temperature, say $T_i = 10^{10}$ GeV, requires looking to equation (4.7) which shows that h_g grows with T_i, and so we expect an even bigger enhancement of the quadrupole.

This discrepancy can be relieved by adjusting τ_i (or $T_i = T_{\rm reheat}$), $h_f(\tau_i)$ or m, since these parameters directly impact the growth in h_g as in (4.7). For example, see figure 4.4 which shows how the bouncing branch power spectrum converges to the standard one as $h_f(\tau_i)/h_g(\tau_i)$ is decreased. Note that these plots would look the same if instead m or T_i was decreased by the same amount. In summary, to achieve a CMB Tensor Power Spectrum that resembles the result from GR, we require very tuned initial conditions, or tuned graviton mass.

4.4 Present Day Stochastic Gravitational Wave Background

We now use the results from the previous section to see how the bigravity primordial gravitational waves contribute to the present day stochastic gravitational wave energy density . The observable quantity of interest is the gravitational wave energy density, defined as a function of frequency:

$$\Omega_{\rm GW}^0(f) = \frac{1}{\rho_{\rm crit}} \frac{d\rho_{\rm GW}}{d\ln f} \tag{4.9}$$

where the critical density is $\rho_{\rm crit} = 3M_{\rm g}^2 H^2(\tau)$ (recall eq. (1.48)).

Direct detection of relic gravitational waves is of considerable interest given the improving technology of ground and space based laser interferometers. Various experiments have already placed upper bounds on $\Omega_{\rm GW}^0$, and proposed experiments will be able to reach much higher sensitivities. Therefore, one might ask if gravitational waves in bigravity would be more or less likely to detect, and if the current sensitives of LIGO or Pulsar Timing Arrays could constrain bigravity. LIGO has already made measurements between $51 < f < 150$ Hz to constrain $\Omega_{\rm GW}^0 < 6.5 \times 10^{-5}$ at these frequencies and advanced LIGO is predicted to reach down to sensitivities of $\Omega_{\rm GW}^0 \sim 6.5 \times 10^{-9}$ in the coming years [225]. In addition, Pulsar-

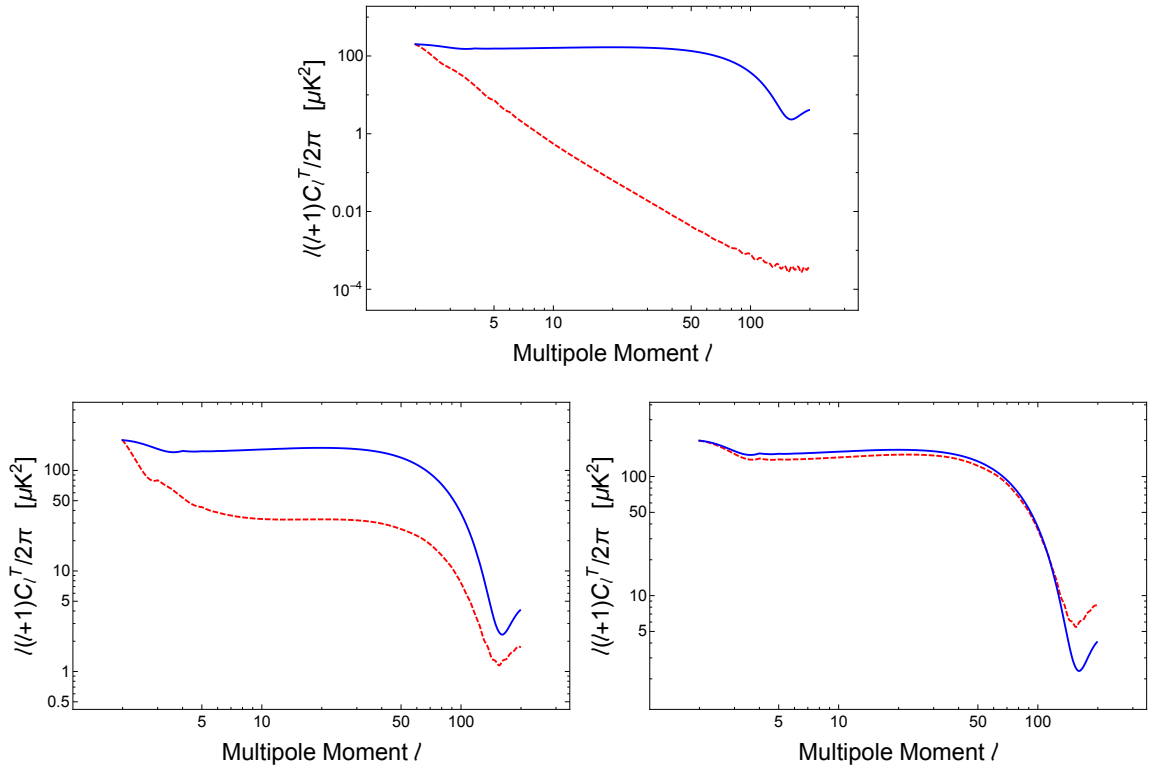

Figure 4.4: The CMB Tensor Power Spectrum in the bouncing branch (red, dashed) approaches the GR result (blue, solid) as the initial value of h_f (or m or T_i) is decreased from left to right: $h_f(\tau_i)/h_g(\tau_i) = 10^{-3},\ 10^{-5},\ 10^{-7}$ with $\tau_i = 10^{-6}H_0^{-1}$. In each cases we have scaled the power spectrum in the bouncing branch down so that $C_{2\ \mathrm{BB}}^T = C_{2\ \mathrm{GR}}^T \sim 200\mu K^2$, which required dividing the spectrum by $\sim 10^6,\ 5.3,\ 1.1$ from left to right.

timing experiments have placed an upper bound of $\Omega_{\mathrm{GW}}^0 < 1.6 \times 10^{-9}$ at low frequencies $10^{-9} < f < 10^{-8}$ Hz [226] and will improve in the future. The first-generation space based laser interferometer, LISA, is expected to operate at sensitivities of $\Omega_{\mathrm{GW}}^0 \sim 10^{-11}$ at frequencies $f \sim 10^{-3}$ Hz [227], while the second-generation space based interferometer, BBO, may be able to reach all the way down to $\Omega_{\mathrm{GW}}^0 \sim 10^{-17}$ near frequencies $f \sim 0.3$ Hz [228].

In terms of the tensor modes (corresponding to the physical metric) the predicted stochastic background can be computed at any conformal time τ via the formula [229]

$$\Omega_{\mathrm{GW}}(k,\tau) = \frac{k^2|h_g(k,\tau)|^2 + |\dot{h}_g(k,\tau)|^2}{12\pi^2\mathcal{H}^2(\tau)} \tag{4.10}$$

given as a function of $k = 2\pi f$. When written with superscript 0, it is understood to be evaluated today $\tau = \tau_0$. More generally however, $\Omega_{\mathrm{GW}}^0(k)$ represents the present-day gravitational wave energy density on scales that re-entered the Hubble horizon during the radiation dominated era.

Over the range of frequencies of interest for the experiments above, the stochastic background due to primordial tensor modes in GR is essentially flat $\Omega_{\mathrm{GW}}^0(k) \sim 10^{-15}$ [230]. The

precise profile depends on the assumed model of inflation that produced the modes, which for us is unimportant as we are just looking for a rough comparison to bigravity.

Let us compare $\Omega^0_{\rm GW}(k)$ in bigravity and GR. In the bouncing branch, we have observed that the late time growth of tensor modes falls off with the square of the frequency, and therefore, using (4.10), so will $\Omega^0_{\rm GW}(k)$. This makes $\Omega^0_{\rm GW}(k)$ harder to detect in the bouncing branch compared to GR over the range of frequencies of interest for the experiments listed above, $10^{-9} < f < 10^3$ Hz (or $10^9 < k/H_0 < 10^{21}$). In the expanding branch, the decay of the tensor modes closely matches with GR, so in this case we expect a result similar to the standard picture. See figure 4.5 for a plot of the results for $\Omega^0_{\rm GW}$ over a range of frequencies from $k = 10H_0$ to $k = 10^4 H_0$.

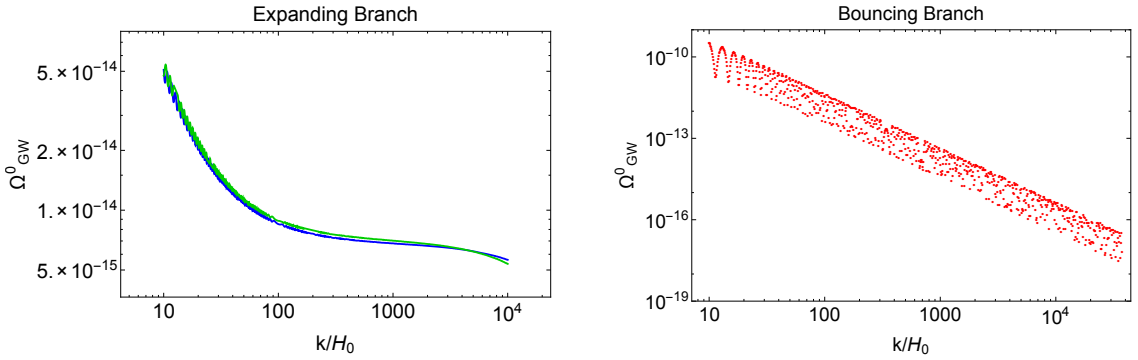

Figure 4.5: The present day stochastic gravitational wave background, given by (4.10) for bigravity as compared to GR. The expanding branch (green) agrees closely with GR (blue), while the bouncing branch (red) shows drastic differences. Note that we have scaled $\Omega^0_{\rm GW}$ down by an appropriate factor so as to fix $C_2^{T\,\rm BB} \sim 200\mu K^2$ (see figure 4.4).

As in the previous section, we see that an adjustment of the initial condition for h_f causes the result for the bouncing branch to converge to the solution in GR, as displayed in figure 4.6. This is equivalent to varying m or T_i by the same amount, as discussed previously.

4.5 Initial Conditions

We have seen that in the bouncing branch, the extreme growth in the dark sector causes amplification of the physical tensor mode, leading to large discrepancies with GR. This amplification causes alterations in physical observables, such as the CMB Power Spectrum and the present day stochastic gravitational wave background. However, if the tensor modes in the dark sector are sufficiently suppressed, then the physical tensor modes and their associated observables closely resemble those of GR. If some mechanism were to exist so that

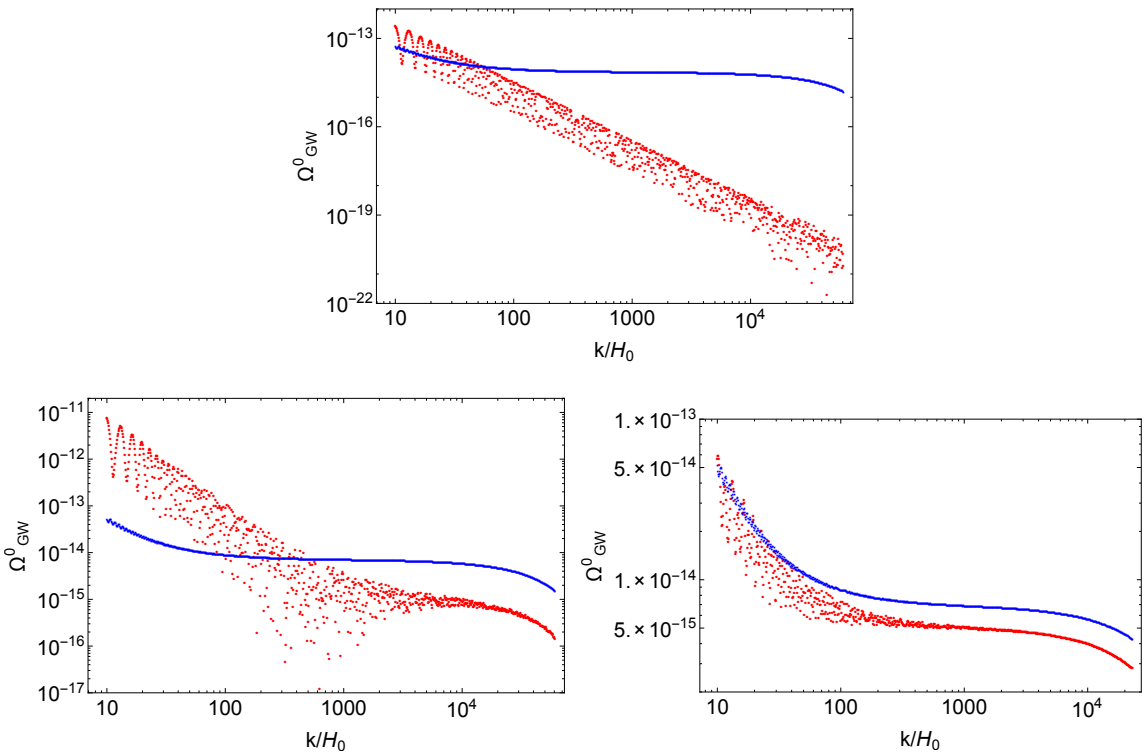

Figure 4.6: The present day stochastic gravitational wave background in the bouncing branch (red) approaches the GR result (blue) as the initial value of h_f (or m or T_i) is decreased from left to right: $h_f(\tau_i)/h_g(\tau_i) = 10^{-3},\ 10^{-5},\ 10^{-7}$ with $\tau_i = 10^{-6}H_0^{-1}$

$h_f(\tau_i) \ll h_g(\tau_i)$ then this branch would have essentially the same gravitational wave spectrum as GR, and would be indistinguishable on an observational level. It is therefore necessary to explore the initial conditions for the primordial tensor modes, assuming they were produced by inflation.

During inflation, the universe undergoes accelerated expansion in a quasi-de Sitter phase. For pure de Sitter, the bigravity background equations simplify as follows:

$$\rho = \text{constant} = 3(H_g^I M_g)^2 \quad \Rightarrow \quad \bar{\rho} = 3(H_g^I/H_0)^2 \tag{4.11}$$

$$r = \text{constant} \tag{4.12}$$

$$c = 1 \quad \Rightarrow \quad \mathcal{H} = \mathcal{H}_g = \mathcal{H}_f \tag{4.13}$$

where H_g^I is the Hubble parameter during inflation. The second line follows from equation (2.49) and the third line follows from (2.40). Therefore, the dark universe is also undergoing de Sitter expansion. Restoring m and H_0 in (2.49) using (2.41), specializing to the bouncing branch parameters in which only β_1^* and β_4^* are nonzero, we get a polynomial equation for the value of the ratio of the scale factors during inflation, r^I:

$$-3(H_g^I)^2 r^I + m^2\left[\beta_1^* + \beta_4^*(r^I)^3 - 3\beta_1^*(r^I)^2\right] = 0. \tag{4.14}$$

Since one typically takes the mass term m in bigravity to be on the order of the Hubble constant H_0 which is much smaller than the Hubble parameter during inflation H_g^I, we can expand the solution for r^I in powers of $H_g^I/m \gg 1$, yielding

$$r^I = \sqrt{\frac{3}{\beta_4^*}}\frac{H_g^I}{m} + \frac{3\beta_1^*}{2\beta_4^*} + \mathcal{O}\left(\frac{m}{H_g^I}\right)^2 \tag{4.15}$$

$$\sim \frac{H_g^I}{H_0} \tag{4.16}$$

where we have used $m = H_0 \Rightarrow \beta_4^* = \beta_4 = 0.94$. Assuming high scale inflation, the maximum allowed Hubble during inflation is $H_g^I \sim 10^{15}$ GeV, and $H_0 \sim 10^{-33}$ eV, from which we can estimate $r^I \sim 10^{57}$. For very low scale inflation, say at the TeV scale, we can estimate $r^I \sim 10^{29}$.

After inflation, there must be a transition to a radiation dominated phase. During the inflationary de Sitter phase, the scale factor for the dark sector metric is increasing, $\dot{b} > 0$, but in the radiation dominated era in the bouncing branch, the scale factor is decreasing, $\dot{b} < 0$. It is evident that the dark sector must undergo another "bounce" transition after inflation from expansion to contraction in order to achieve the necessary behaviour in the early radiation era. In the next section we see that this bounce is indeed achieved in a simple inflationary model.

A note on inflation in the expanding branch is now in order. In this branch, we set $\beta_4^* = 0$. Solving for r^I in (4.14) with only $\beta_1^* \neq 0$, we obtain

$$r^I = \frac{-3{H_g^I}^2 + \sqrt{9{H_g^I}^2 + 12m^4{\beta_1^*}^2}}{6m^2\beta_1^*} \tag{4.17}$$

$$= \frac{\beta_1^* m^2}{3{H_g^I}^2} + \mathcal{O}\left(\frac{m}{H_g^I}\right)^3 \tag{4.18}$$

$$\sim \left(\frac{H_0}{H_g^I}\right)^2 \tag{4.19}$$

Therefore, in the expanding branch we obtain a very small value of r^I.

To find the power spectrum of primordial tensors, we expand the action eq. 2.37 to quadratic order in transverse traceless perturbations of the f and g metrics. During inflation, the interaction terms are unimportant due to the large hierarchy between m and H_g^I. Defining the canonically normalized fields:

$$v_g = \frac{M_g a}{2}h_g, \quad v_f = \frac{M_g b}{2}h_f, \tag{4.20}$$

and using the fact that

$$\frac{\ddot{a}}{a} = \frac{\ddot{b}}{b} = \frac{2}{\tau^2} \tag{4.21}$$

in de Sitter, the quadratic action for tensors during an inflationary epoch in bigravity is given by:

$$S = \sum_{+,\times} \frac{1}{2} \int d\tau \; d^3k \left[\dot{v}_{g,f}^2 - \left(k^2 - \frac{2}{\tau^2} \right) v_{g,f}^2 \right] \tag{4.22}$$

Imposing Bunch Davies initial conditions [231], the mode functions each obey:

$$v_{g,f} = \frac{1}{k^{3/2}\tau} \left(1 - ik\tau\right) e^{ik\tau} \tag{4.23}$$

Summing over the two polarization states, the power spectrum for h_f and h_g are given by:

$$\Delta_T^{(g)^2} = \frac{2k^3}{\pi^2 M_g^2} \frac{|v_g|^2}{a^2}\bigg|_{k=\hat{\mathcal{H}}}, \quad \Delta_T^{(f)^2} = \frac{2k^3}{\pi^2 M_g^2} \frac{|v_f|^2}{b^2}\bigg|_{k=\hat{\mathcal{H}}} \tag{4.24}$$

Substituting with Eqn. 4.23 and using $H_g^I = (a\tau)^{-1}$ for a de Sitter phase, we obtain:

$$\Delta_T^{(g)^2} = \frac{2H_g^{I^2}}{\pi^2 M_g^2}\bigg|_{k=\hat{\mathcal{H}}}, \quad \Delta_T^{(f)^2} = \frac{\Delta_T^{(g)^2}}{r^{I^2}} \tag{4.25}$$

Modes are populated for both the visible and dark sector tensors on all super horizon scales. The appearance of r^I is a consequence of the fact that the relative size of a and b is physical, and cannot be removed by a change of coordinates or field redefinition.

In the bouncing branch $r^I \gg 1$, leading to a drastic suppression in the initial amplitude for h_f roughly given by $h_f/h_g = 1/r^I \sim H_0/H_g^I$. This suppression is more than sufficient to bring the amplitude of dark sector tensor fluctuations below the threshold where they alter the propagation of visible sector tensors. With inflationary initial conditions, we therefore conclude that there would be no visible deviation from GR in the tensor contribution to the CMB temperature anisotropies or the late time stochastic distribution of gravitational waves. This is true for both high scale and low scale inflation.

In contrast, the expanding branch gives $h_f/h_g \sim (H_g^I/H_0)^2$, which for any reasonable choice of the inflationary scale is far beyond the perturbative regime. Therefore, we cannot make sense of inflation in the expanding branch with $\beta_4 = 0$. If instead we considered a non-minimal expanding branch with $\beta_4 \neq 0$, then the result would be the same as the bouncing branch.

Finally, let us comment on the time dependence of the dark sector tensor modes. This is of interest because as illustrated in Cusin et. al. [216], the dark sector tensor mode has a growing mode on super horizon scales proportional to τ^3 during radiation domination. We can estimate the relative amplitude of the growing and constant modes at the beginning of radiation domination as

$$\frac{\tau_i \dot{h}_f(\tau_i)}{h_f(\tau_i)} = \frac{k^2 \tau_i}{a(\tau_i) H_g^I} \tag{4.26}$$

Using $a \propto \tau$ during radiation domination, for horizon-scale wave numbers, in the bouncing branch we can estimate this ratio as $\sim H_0/H_g^I \sim 10^{-57}$. For any reasonable choice of reheat temperature, this will be smaller than the growth factor for the growing mode during radiation domination, $\sim (\tau_{\rm eq}/\tau_i)^3 = (T_{\rm reh}/T_{\rm eq})^3 \sim 10^{30}$, where the subscript "eq" refers to matter radiation equality. We therefore conclude that the appropriate initial conditions are $\dot{h}_{g,f} = 0$.

4.5.1 An inflationary model in bigravity

In this section, we examine bigravity for the $m^2\varphi^2$ inflationary model originally presented in section 1.3.2. We want to determine the behaviour of r (and thus b) during inflation in the bouncing branch with only $\beta_{1,4} \neq 0$. The inflaton field has potential energy $V(\varphi) = \frac{1}{2}m_\varphi^2\varphi^2$ and energy density $\rho_\varphi = \frac{1}{2}(\partial_t\varphi)^2 + V(\varphi)$. For this calculation we find it convenient to use the following set of dimensionless variables:

$$\tilde{H} = \frac{H}{H_g^I} \qquad \tilde{t} = tH_g^I \qquad \tilde{\Gamma}_\varphi = \frac{\Gamma_\varphi}{H_g^I} \qquad \tilde{V} = \frac{V}{M_g^2(H_g^I)^2} \tag{4.27}$$

$$\tilde{\varphi} = \frac{\varphi}{M_g} \qquad \tilde{\beta}_n = \frac{m^2}{(H_g^I)^2}\beta_n^* = \frac{H_0^2}{(H_g^I)^2}\beta_n \qquad \tilde{\rho}_r = \frac{\rho_r}{M_g^2(H_g^I)^2} \tag{4.28}$$

where Γ_φ is the decay rate of the inflaton, t is proper time, and we take $(H_g^I)^2 = V(\varphi_0)/3M_g^2$ in terms of the value of φ at the start of inflation, implying that $\tilde{V} = 3\tilde{\varphi}^2/\tilde{\varphi}_0^2$.

Recalling equation (1.67), and including an explicit decay of the inflaton into radiation, the equation of motion is

$$\tilde{\varphi}'' + 3\tilde{H}\tilde{\varphi}' + \tilde{\Gamma_\varphi}\tilde{\varphi}' + 6\frac{\tilde{\varphi}}{\tilde{\varphi_0}^2} = 0 \tag{4.29}$$

where a prime denotes a derivative with respect to dimensionless proper time $\tilde{t}$. The Friedmann equation (2.42) for $a(\tilde{t})$ becomes

$$3\tilde{H}^2 = \tilde{\rho}_\varphi + \tilde{\rho}_r = \frac{1}{2}(\tilde{\varphi}')^2 + 3\frac{\tilde{\varphi}^2}{\tilde{\varphi}_0^2} + \tilde{\rho}_r. \tag{4.30}$$

Notice that we have neglected the contribution of $\tilde{\rho}_m$ and $\tilde{\rho}_{\rm mg}$ in the early universe since these will be highly suppressed compared to the inflaton or radiation energy density. The radiation energy density $\tilde{\rho}_r$ satisfies a modified conservation equation

$$\tilde{\rho}_r' + 4\tilde{H}\tilde{\rho}_r = \tilde{\Gamma}_\varphi\tilde{\rho}_\varphi \tag{4.31}$$

We can now solve (4.29), (4.30), and (4.31) for the functions $\tilde{\varphi}(\tilde{t})$, $a(\tilde{t})$, $\tilde{\rho}_r(\tilde{t})$. The last ingredient will be to solve for $r(\tilde{t})$, for which we use (2.49), which simplifies in the bouncing branch with $\beta_{1,4} \neq 0$ to

$$0 = r\tilde{\rho} - \tilde{\beta}_1 - \tilde{\beta}_4 r^3 + 3\tilde{\beta}_1 r^2 \tag{4.32}$$

$$r = \frac{H_g^I}{H_0}\sqrt{\frac{\tilde{\rho}}{\beta_4}} \quad \text{for large } r \tag{4.33}$$

The results of the calculation are shown in figure 4.7. We can see the transition from $b' > 0$ during inflation to $b' < 0$ after inflation is achieved in the bottom right plot. Notice that after b hits its first peak, it oscillates as it decreases, indicating brief periods of expansion and contraction of $f_{\mu\nu}$. This behaviour is caused by the oscillation of the inflaton around its minimum, and implies that the dark sector metric undergoes multiple bounces during reheating.

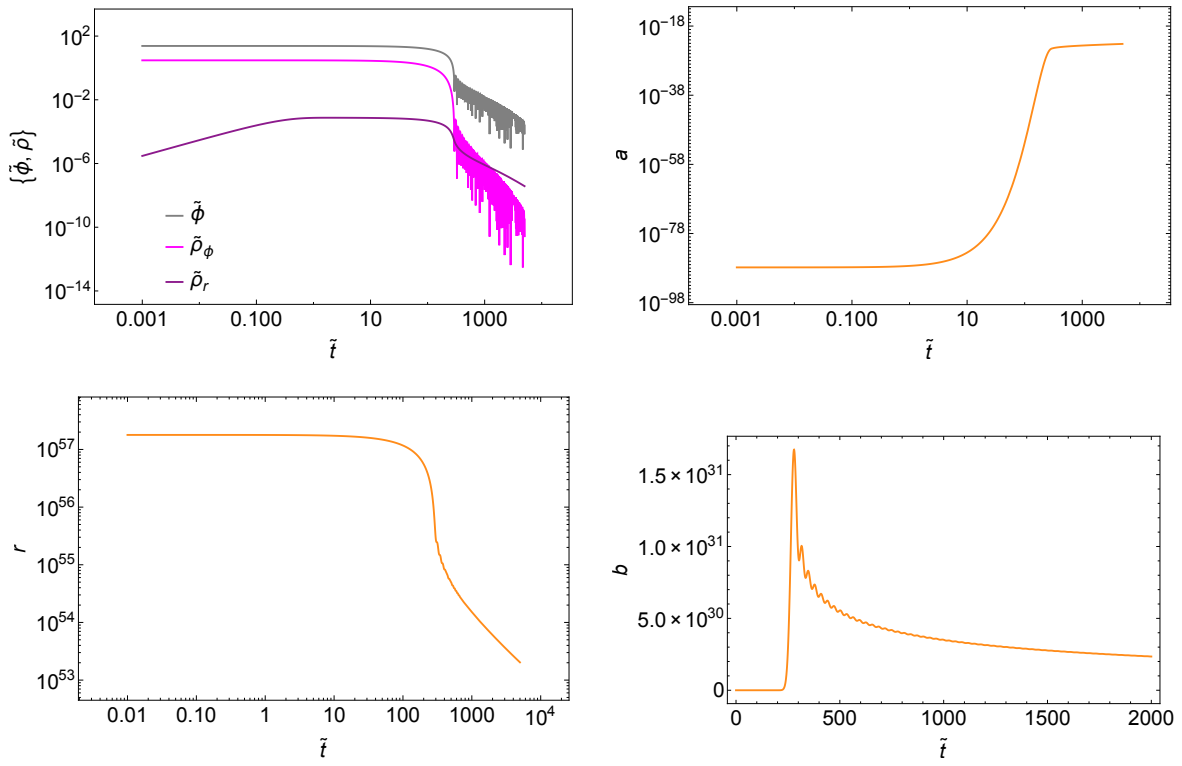

Figure 4.7: Top left: The solutions of (4.29), (4.30), and (4.31) for $\tilde{\varphi}$, $\tilde{\rho}_{\varphi}$, and $\tilde{\rho}_r$ vs. $\tilde{t}$ for the $m^2\varphi^2$ model of inflation. We set $\Gamma_{\varphi} = 10^{-3}$, $\tilde{\varphi}(0) = \tilde{\varphi}_0 = 24$, $\tilde{\varphi}'(0) = 0$, and $\tilde{\rho}_r(0) = 0$. Top right: The scale factor for $g_{\mu\nu}$. Bottom left: The ratio r found via (4.33). Bottom right: The scale factor for $f_{\mu\nu}$, $b = ra$.

4.6 Conclusion

We have studied the properties of gravitational waves in massive bigravity, and their impact on cosmological observables compared to the standard predictions of General Relativity. The two background solutions we have studied display dramatically different phenomenology, illustrating the enormous size of the parameter space for observables.

In the "expanding branch" in which both metrics expand in time, we found that for a wide range of initial conditions, the physical tensor perturbation h_g matches closely with the pure GR solution. Due to a dramatic decay of h_f, the impact of the dark sector on h_g is not

important and causes no significant deviation from GR.

The "bouncing branch", in which the dark metric $f_{\mu\nu}$ undergoes a bounce from contraction to expansion, has potentially dramatic differences from GR in the tensor sector. When the f metric is undergoing contraction, the lapse c is negative, which causes h_f to grow. This growth in h_f can translate into growth in h_g through the mixing term in the equations of motion, in some cases leading to physical gravitational waves with growing amplitudes at late times. This contrasts significantly with gravitational waves in GR which decay with time. The growth can potentially impact the CMB tensor power spectrum by dramatically amplifying large scale temperature anisotropies. The present day stochastic gravitational wave background, $\Omega^0_{\rm GW}$ can be impacted as well, inheriting a very red spectrum that decays with the square of the wave number. The degree of growth depends on the scale of reheating, amplitude of the dark sector tensor modes, wave number, and graviton mass, and obeys the scaling relation Eq. 4.7 for initial amplitudes above a critical value. On the largest scales, we find that the dark sector tensor modes can have a significant influence on the physical tensor for $h_f(\tau_i)/h_g(\tau_i) > 10^{-20}$ for a reheat temperature of $T_i = 10^{10}$ GeV. In the absence of a theory of initial conditions, it is not clear that this holds.

To address the question of initial conditions, we computed the primordial power spectrum for dark and visible sector tensors in an inflationary cosmology. We found that the expanding branch is far beyond the perturbative regime, and therefore inaccessible to a semi-classical treatment. However, the primordial power spectra in the bouncing branch show that $h_f/h_g \sim H_0/H_g^I \sim 10^{-57}$ for high scale inflation, and $h_f/h_g \sim 10^{-29}$ for low scale inflation. With this level of suppression, there will be no observable deviation from GR in the CMB or stochastic gravitational wave background. We presented an inflationary model that exhibits this explicitly.

Let us now briefly discuss our results in the context of previous work in Refs. [216, 217]. These authors considered the bouncing branch chosen in this work with identical parameters, but with an initial condition that was entirely composed of a growing mode of the dark sector tensor. However, as shown in section 4.5 it appears that inflationary initial conditions do not excite the growing mode as significantly. Of course, another theory of initial conditions may prevail, and a proper treatment of higher-order couplings between the scalar and tensor sectors may reveal a significant enhancement. In this sense, our investigation is largely complementary to previous work; taken together, the range of possibilities is covered. However, in the context of inflationary cosmology, it appears that the growing mode is not excited at the end of inflation.

It is clear that tensors can be a sensitive probe of massive bigravity. Looking to the future, the parameter space of nearly homogeneous solutions will soon be completely explored both

at the level of the background and first order perturbations. In light of this, it is equally important to consider the theory for initial conditions in a broader sense, as illustrated by the strong dependence on initial conditions found in this and other papers. To this end, we plan to return to the question of inflationary model building in massive bigravity in future work. Scenarios with small but observable deviations from GR could serve as an important alternative hypothesis necessary for testing GR on cosmological scales and in future gravitational wave observatories.

Chapter 5

Reducing cosmic variance with the large-scale kSZ effect

Due to cosmic variance we cannot learn any more about large-scale inhomogeneities from the primary cosmic microwave background (CMB) alone. More information on large scales is essential for resolving large angular scale anomalies in the CMB. Here we consider cross correlating the large-scale kinetic Sunyaev Zel'dovich (kSZ) effect and probes of large-scale structure, a technique known as kSZ tomography. The statistically anisotropic component of the cross correlation encodes the CMB dipole as seen by free electrons throughout the observable universe, providing information about long wavelength inhomogeneities. We compute the large angular scale power asymmetry, constructing the appropriate transfer functions, and estimate the cosmic variance limited signal to noise for a variety of redshift bin configurations. The signal to noise is significant over a large range of power multipoles and numbers of bins. We present a simple mode counting argument indicating that kSZ tomography can be used to estimate more modes than the primary CMB on comparable scales. A basic forecast indicates that a first detection could be made with next-generation CMB experiments and galaxy surveys. This chapter motivates a more systematic investigation of how close to the cosmic variance limit it will be possible to get with future observations.

5.1 Introduction

The CMB has been an extraordinarily powerful tool for precision cosmology, establishing the standard model, ΛCDM, at high confidence. However, on very large-scales, CMB measurements are limited by cosmic variance, implying that we can not hope to learn any more about large scale inhomogeneities from the primary CMB alone. As introduced in section 1.6, cosmic variance is an inherent sample variance that arises because we only have one observable universe to measure. This limiting precision restricts our ability to discriminate different models. For instance, more information on large scales is essential to distinguish between various infrared modifications of GR.

How can we overcome cosmic variance and unlock the secrets of our universe on large scales? In attempting to address this and other issues, the field of cosmology is now transitioning away from the primary CMB and towards understanding large scale structure (LSS). The three dimensional nature of LSS means that it contains more information than the two dimensional last scattering surface of the CMB. Constraints from probes of LSS, such as next-generation galaxy surveys (e.g [232, 233]) and 21cm measurements (e.g. [234, 235]), are poised to become increasingly important for many cosmological parameters. Even still, when it comes to measuring inhomogeneities on scales $\gtrsim$ Gpc, there will be limited additional constraining power in all but the most ambitious scenarios (e.g. 21cm dark ages cosmology [236]).

In this project [3], we investigate the viability of an additional probe of large scale inhomogeneities: large scale kinetic Sunyaev Zel'dovich (kSZ) tomography (introduced at the end of section 1.6). The kSZ effect is a CMB temperature anisotropy arising from the Compton scattering of CMB photons by the bulk motion of free electrons with respect to the CMB rest frame [80]. The contribution to the kSZ effect from each free electron is proportional to the locally observed CMB dipole, and because each free electron probes a different portion of the surface of last scattering, measurements of the kSZ effect can in principle yield information about the homogeneity of the universe. This is why measurements of the kSZ effect can be so constraining for scenarios that predict a deviation from large-scale homogeneity. Note that this is a dramatically different regime than the one typically explored, for example in the pairwise motion of clusters that yielded the first detection. This large-scale kSZ effect is sensitive to the Sachs Wolfe and integrated Sachs Wolfe components of the dipole, in addition to the Doppler component from peculiar velocities, and can therefore in principle yield more large-scale information than peculiar velocity surveys or direct measurements of the density field through various tracers of LSS.

Cross correlation of the kSZ effect with tracers of LSS, known as kSZ tomography, is key

to extracting the most information possible. In this project we will show that information about large scale homogeneity is encoded in a statistical anisotropy of the direct correlation of tracers of large scale structure and the small angular scale CMB, e.g. a power asymmetry. Importantly, the contribution to this signal from small-scale peculiar velocities vanishes [88].

The power of kSZ tomography to probe the large scale homogeneity of the universe has been highlighted previously, notably in refs. [101, 89, 112, 113]. These papers considered theoretical extensions to ΛCDM where a signal could hopefully be detected with current and near-future experiments. However, as the sensitivity and resolution of CMB experiments continues to develop and as our ability to probe LSS improves, we might hope to enter an era where large scale kSZ tomography becomes a tool not just for constraining exotic scenarios, but for measuring the inhomogeneities we know to exist: those responsible for the large scale temperature anisotropies in the primary CMB.

The aim of this project is to explore this eventuality in the most optimistic, cosmic variance limited, scenario. More specifically, we compute the angular spectrum of the asymmetry in the kSZ-LSS cross power expected in ΛCDM as a function of redshift. Comparing this signal to the accidental power asymmetry expected from the statistically isotropic components of the kSZ effect (the dominant source of CMB temperature anisotropies on small angular scales), we find that the signal-to-noise can be significant ($S/N \sim \mathcal{O}(10^2 - 10^3)$) over a wide range of angular scales ($\ell_{\rm max} \sim \mathcal{O}(100)$) and in a large number ($N_{\rm bin} \sim \mathcal{O}(10 - 100)$) of redshift bins. A simple mode counting argument indicates that there is in principle more information in the power asymmetry than in the primary CMB on the relevant scales, for a sufficient number of redshift bins ($N_{\rm bin} \gtrsim 30$). We present a basic forecast, indicating that a first detection could be made with next-generation CMB experiments and galaxy surveys.

The plan of the chapter is as follows. In section 5.2, the large-scale late-time kSZ effect is summarized along with a derivation of the large-scale effective velocity. Section 5.3 describes simulations of the large-scale effective velocity field. Section 5.4 outlines how kSZ tomography can be used to extract the large-scale effective velocity. Then, we derive the cosmic variance limited noise in section 5.5 and estimate the signal using both an RMS estimate and simulations in section 5.6. In section 5.7 we provide an estimate for the number of modes that can be obtained using cosmic variance limited kSZ tomography, showing that in principle more information can be extracted than is contained in the primary CMB on comparable scales. Finally, we assess the detectability of the signal with next-generation CMB experiments and galaxy surveys in section 5.8; we conclude in section 5.9. A number of results are collected in the Appendices.

5.2 The large-scale kSZ effect

The kinetic Sunyaev Zel'dolvich effect arises from Compton scattering of CMB photons by free electrons moving with respect to the CMB rest frame. This produces temperature anisotropies given by an integral along the line of sight:

$$\left.\frac{\Delta T}{T}\right|_{\rm kSZ}(\hat{\mathbf{n}}_e) = -\sigma_T \int_0^{\chi_{\rm re}} d\chi_e \; a_e(\chi_e) \; n_e(\hat{\mathbf{n}}_e, \chi_e) \; \mathbf{v}_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) \cdot \hat{\mathbf{n}}_e \tag{5.1}$$

$$= -\sigma_T \int_0^{\chi_{\rm re}} d\chi_e \; a_e(\chi_e) \; \bar{n}_e(\chi_e) \; (1 + \delta(\hat{\mathbf{n}}_e, \chi_e)) \; v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e). \tag{5.2}$$

The geometry is depicted in figure 5.1. In eq. (5.1), σ_T is the Thomson cross-section, $n_e(\hat{\mathbf{n}}_e, \chi_e)$ is the electron number density, $\hat{\mathbf{n}}_e$ is the angular direction on the sky to the scatterer, and χ_e is the comoving radial coordinate to the scatterer along our past light cone (recall eq. (1.53)),

$$\chi_e = \int_0^{z_e} \frac{dz}{H(z)} = -\int_1^{a_e} \frac{da}{H(a)a^2}, \tag{5.3}$$

where z_e and a_e are the scatterer's redshift and scale factor respectively. Below, we will use χ_e and z_e interchangeably. In the second line of (5.2), we have written the electron number density as $n_e(\hat{\mathbf{n}}_e, \chi_e) = \bar{n}_e(\chi_e)(1 + \delta(\hat{\mathbf{n}}_e, \chi_e))$ in terms of the average electron number density $\bar{n}_e(\chi_e)$, and the density perturbation δ (recall eq. (1.73)), and replaced $\mathbf{v}_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) \cdot \hat{\mathbf{n}}_e$ with the projection along the line of sight $v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e)$.

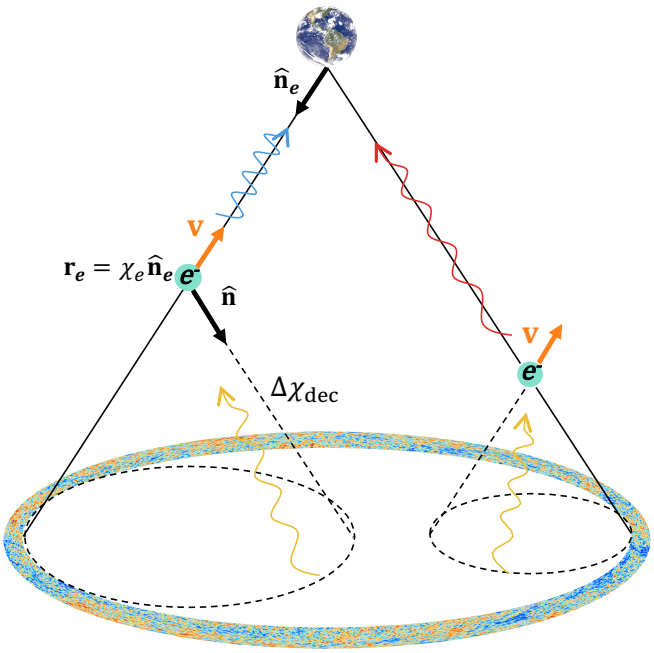

Figure 5.1: Scattering of CMB photons off free electrons on our past light cone. The position of an electron is described in terms of its direction $\hat{\mathbf{n}}_e$ and comoving distance χ_e. The direction from the electron to a point on the surface of last scattering is denoted by $\hat{\mathbf{n}}$ and the distance to last scattering by $\Delta\chi_{\rm dec}$.

The quantity denoted by $v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e)$ is the CMB dipole observed by each electron, projected along the line of sight:

$$v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) = \frac{3}{4\pi} \int d^2\hat{\mathbf{n}}\ \Theta(\hat{\mathbf{n}}_e, \chi_e, \hat{\mathbf{n}})\ (\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_e), \tag{5.4}$$

where, for a freely falling electron at position $\mathbf{r}_e \equiv \chi_e \hat{\mathbf{n}}_e$, the CMB temperature it sees along the direction $\hat{\mathbf{n}}$ is given by

$$\Theta(\hat{\mathbf{n}}_e, \chi_e, \hat{\mathbf{n}}) = \Theta_{\rm SW}(\hat{\mathbf{n}}_e, \chi_e, \hat{\mathbf{n}}) + \Theta_{\rm Doppler}(\hat{\mathbf{n}}_e, \chi_e, \hat{\mathbf{n}}) + \Theta_{\rm ISW}(\hat{\mathbf{n}}_e, \chi_e, \hat{\mathbf{n}}). \tag{5.5}$$

The three contributions come from the Sachs-Wolfe (SW) effect generated by the gravitational potential on the LSS, the Doppler effect due to peculiar motion of photons on the LSS and peculiar motion of electrons at redshift z_e, and the integrated Sachs-Wolfe (ISW) effect.

Working in Newtonian gauge

$$ds^2 = -(1+2\Psi)dt^2 + a^2(t)(1-2\Psi)d\mathbf{x}^2, \tag{5.6}$$

the Sachs-Wolfe contribution is given by

$$\Theta_{\rm SW}(\hat{\mathbf{n}}_e, \chi_e, \hat{\mathbf{n}}) = \left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2}\right) \Psi_i(\mathbf{r}_{\rm dec}), \tag{5.7}$$

where $\mathbf{r}_{\rm dec} \equiv \chi_e \hat{\mathbf{n}}_e + \Delta\chi_{\rm dec}\hat{\mathbf{n}}$ with $\Delta\chi_{\rm dec} = \Delta\chi(a_{\rm dec}) = -\int_{a_e}^{a_{\rm dec}} da \left(H(a)a^2\right)^{-1}$ the distance along the electron's past light cone to decoupling. More generally, we will define

$$\Delta\chi(a) = -\int_{a_e}^{a} \frac{da}{H(a)a^2}. \tag{5.8}$$

In eq. (5.7) we have used the growth function, $D_\Psi(a)$, which relates the potential to its primordial value at $a \to 0$ through the definition $\Psi(\mathbf{r}, a) = D_\Psi(a)\Psi_i(\mathbf{r})$. The growth function is well approximated on superhorizon scales by

$$D_\Psi(a) \equiv \frac{\Psi_{\rm SH}(a)}{\Psi_{\rm SH,i}} = \frac{16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16}{10y^3} \left[\frac{5}{2}\Omega_m \frac{E(a)}{a} \int_0^a \frac{da}{E^3(a)\ a^3}\right], \tag{5.9}$$

where $y \equiv a/a_{\rm eq}$ and $E(a) = \sqrt{\Omega_m a^{-3} + \Omega_\Lambda}$ is the normalized Hubble parameter.

The Doppler component is given by

$$\Theta_{\rm Doppler}(\hat{\mathbf{n}}_e, \chi_e, \hat{\mathbf{n}}) = \hat{\mathbf{n}} \cdot [\mathbf{v}(\mathbf{r}_e, \chi_e) - \mathbf{v}(\mathbf{r}_{\rm dec}, \chi_{\rm dec})]. \tag{5.10}$$

The velocities can be related to the potential through

$$\mathbf{v} = -\frac{2a^2c^2H(a)}{H_0^2\Omega_m} \frac{y}{4+3y} \left[\nabla\Psi + \frac{d\nabla\Psi}{d\ln a}\right], \tag{5.11}$$

which is valid on all scales. On large scales, we can use this expression to define a velocity growth function $D_v(a)$:

$$\mathbf{v} = -\frac{2a^2c^2H(a)}{H_0^2\Omega_m} \frac{y}{4+3y} \left[D_\Psi + \frac{dD_\Psi}{d\ln a}\right] \nabla\Psi_i, \tag{5.12}$$

where

$$D_v(a) \equiv \frac{2a^2 H(a)}{H_0^2 \Omega_m} \frac{y}{4+3y} \left[D_\Psi + \frac{dD_\Psi}{d\ln a} \right]. \tag{5.13}$$

Finally, the ISW term is given by

$$\Theta_{\rm ISW}(\hat{\mathbf{n}}_e, \chi_e, \hat{\mathbf{n}}) = 2 \int_{a_{\rm dec}}^{a_e} \frac{d\Psi}{da}(\mathbf{r}(a), a) da = 2 \int_{a_{\rm dec}}^{a_e} \frac{dD_\Psi}{da} \Psi_i(\mathbf{r}(a)) da. \tag{5.14}$$

Here, $\mathbf{r}(a) = \mathbf{r}_e + \Delta\chi(a)\ \hat{\mathbf{n}}$.

5.2.1 Fourier kernel for the effective velocity

Relating each contribution to the effective velocity to the primordial potential Ψ_i allows us to define a kernel in Fourier space relating Ψ_i to the effective velocity. The details can be found in Appendix C, which results in the expression

$$v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) = i \int \frac{d^3k}{(2\pi)^3}\ T(k) \tilde{\Psi}_i(\mathbf{k})\ \mathcal{K}^v(k, \chi_e)\ \mathcal{P}_1(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}}_e)\ e^{i\chi_e \mathbf{k}\cdot\hat{\mathbf{n}}_e}, \tag{5.15}$$

where we have incorporated the transfer function $T(k)$ for the potential to account for sub-horizon evolution on small scales and the kernel $\mathcal{K}^v(k, \chi_e)$ receives contributions from the SW, ISW, and Doppler terms

$$\mathcal{K}^v(k, \chi_e) \equiv \left[\mathcal{K}_{\rm D}(k, \chi_e) + \mathcal{K}_{\rm SW}(k, \chi_e) + \mathcal{K}_{\rm ISW}(k, \chi_e) \right], \tag{5.16}$$

given by

$$\mathcal{K}_{\rm D}(k, \chi_e) \equiv k D_v(\chi_{\rm dec}) j_0(k\Delta\chi_{\rm dec}) - 2k D_v(\chi_{\rm dec}) j_2(k\Delta\chi_{\rm dec}) - k D_v(\chi_e), \tag{5.17}$$

$$\mathcal{K}_{\rm SW}(k, \chi_e) \equiv 3 \left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2} \right) j_1(k\Delta\chi_{\rm dec}), \tag{5.18}$$

$$\mathcal{K}_{\rm ISW}(k, \chi_e) \equiv 6 \int_{a_{\rm dec}}^{a_e} da \frac{dD_\Psi}{da}\ j_1(k\Delta\chi(a)). \tag{5.19}$$

The SW and ISW kernels have support predominantly on large scales, while the Doppler kernel has support on all scales. The last term in the Doppler kernel, $kD_v(\chi_e)$, yields the "conventional" kSZ effect, and represents the dominant contribution to the kSZ effect on scales that have currently been measured (e.g. using pairwise cluster velocities). Using the fact that $j_1(k\Delta\chi(a))$ and $j_2(k\Delta\chi(a))$ have support predominantly on scales $k \sim 1/\Delta\chi$, we can estimate the order of magnitude of scales that contribute to these terms. In the range of redshift between $6 > z_e > 0$, we have $1.25 < H_0\Delta\chi_{\rm dec} < 3.18$. Using $k = 1/\Delta\chi_{\rm dec}$, this translates into the range $(14.3\ {\rm Gpc})^{-1} < k < (5.6\ {\rm Gpc})^{-1}$.

There is one important physical condition that must hold: a pure potential gradient should not contribute to an observable like the kSZ effect [237]. A pure gradient can be removed in

linear perturbation theory by performing a special conformal transformation on the spatial metric. More generally, it is always possible to remove the gradient at a point by the same special conformal transformation. We include a proof of these statements in Appendix D. This absence of a gradient contribution to the kSZ effect has important implications for the behavior of the effective velocity Fourier kernel $\mathcal{K}^v$ at small k.

To see this, consider a Newtonian potential that is a pure gradient

$$\Psi_i(\mathbf{x}) = A_j x^j. \tag{5.20}$$

Using the properties of the derivative of the Dirac delta function, we can write this in Fourier space as

$$\Psi_i(\mathbf{k}) = i(2\pi)^3 A^j \ \partial_j \delta^3(\mathbf{k}). \tag{5.21}$$

Evaluating eq. (5.15), we obtain

$$v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) = -A^j \partial_j \mathcal{K}^v(k=0, \chi_e), \tag{5.22}$$

where we have used the fact that $\mathcal{K}^v(k=0,\chi_e) = 0$, $T(k=0) = 1$, and $\partial_j T(k=0) = 0$. Unless $\partial_j \mathcal{K}^v(k=0,\chi_e) = 0$, there will be an observable kSZ effect from a pure gradient, which would be unphysical. This, together with the fact that each of the three contributions to $\mathcal{K}^v(k,\chi_e)$ are odd functions of k, implies that we must have

$$\mathcal{K}^v(k \to 0, \chi_e) = \mathcal{O}(k^3) + \ldots \tag{5.23}$$

Expanding $\mathcal{K}_{\rm SW}$, $\mathcal{K}_{\rm ISW}$, and $\mathcal{K}_{\rm D}$ separately, the leading order term in the Taylor series expansion is linear in k. Therefore, a cancellation between these terms must occur in the limit $k \to 0$. This is the same type of cancellation demonstrated for the primary CMB dipole due to a pure gradient in Ref. [237]. We demonstrate this cancellation analytically in a Universe with matter and a cosmological constant in Appendix E. The cancellation, and the behavior of the full kernel $\mathcal{K}^v(k,\chi_e)$ at $z=1$ for small k is shown in figure 5.2.

Looking at the inset of the right panel in figure 5.2, we see that the contributions from the spherical Bessel functions in the SW, ISW, and Doppler kernels are evident as oscillations on the linear Doppler contribution. These effects are significant on scales up to of order $k \sim 10\ H_0 \sim 2 \times 10^{-3}\ \mathrm{Mpc}^{-1}$.

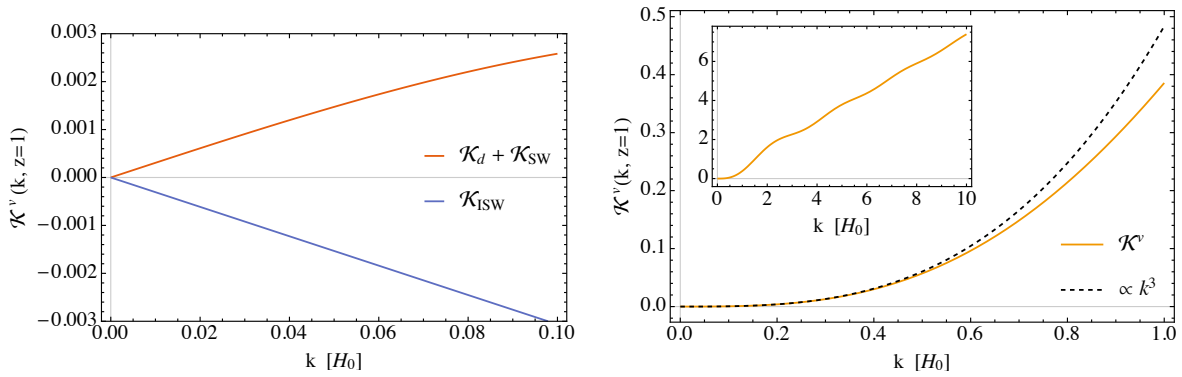

Figure 5.2: The cancellation at linear order of the three pieces of $\mathcal{K}^v$ as $k \to 0$ in ΛCDM. This cancellation can be shown analytically in a universe without radiation (see Appendix E). The leading order behavior is cubic as $k \to 0$ as shown in the right panel.

5.2.2 Angular decomposition of the effective velocity

The multipole moments of the effective velocity are given by integrating over $\hat{\mathbf{n}}_e$

$$a^v_{\ell m}(\chi_e) = \int d^2\hat{\mathbf{n}}_e \; v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) \; Y^*_{\ell m}(\hat{\mathbf{n}}_e) \tag{5.24}$$

$$= i \int \frac{d^3k}{(2\pi)^3} \; T(k) \; \tilde{\Psi}_i(\mathbf{k}) \; \mathcal{K}^v(k, \chi_e) \int d^2\hat{\mathbf{n}}_e \; Y^*_{\ell m}(\hat{\mathbf{n}}_e) \; \mathcal{P}_1(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}}_e) \; e^{i\chi_e \mathbf{k}\cdot\hat{\mathbf{n}}_e}. \tag{5.25}$$

The second integral can be written as a triple product of spherical harmonics by expanding the exponential using (C.3) and writing the Legendre polynomials in terms of spherical harmonics using

$$\mathcal{P}_\ell(\hat{\mathbf{x}} \cdot \hat{\mathbf{x}}') = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\hat{\mathbf{x}}) \; Y_{\ell m}(\hat{\mathbf{x}}'). \tag{5.26}$$

The integral over $\hat{\mathbf{n}}_e$ then becomes

$$\int d^2\hat{\mathbf{n}}_e Y^*_{\ell m}(\hat{\mathbf{n}}_e) \left[\frac{4\pi}{3} \sum_{m''=-1}^{1} Y^*_{1m''}(\hat{\mathbf{k}}) Y_{1m''}(\hat{\mathbf{n}}_e)\right] \left[4\pi i^{\ell'} j_{\ell'}(k\chi_e) \sum_{m'=-\ell'}^{\ell'} Y_{\ell' m'}(\hat{\mathbf{k}}) Y^*_{\ell' m'}(\hat{\mathbf{n}}_e)\right]$$

$$= \frac{(4\pi)^2}{3} \sum_{\ell', m', m''} i^{\ell'} \, j_{\ell'}(k\chi_e) Y^*_{1m''}(\hat{\mathbf{k}}) Y_{\ell' m'}(\hat{\mathbf{k}}) \int d^2\hat{\mathbf{n}}_e Y^*_{\ell m}(\hat{\mathbf{n}}_e) Y^*_{\ell' m'}(\hat{\mathbf{n}}_e) Y_{1m''}(\hat{\mathbf{n}}_e). \tag{5.27}$$

The triple product integral can be expressed in terms of the Clebsch-Gordan coefficients $\mathrm{C}^{\ell_1\ell_2\ell_3}_{m_1 m_2 m_3}$:

$$\int d^2\hat{\mathbf{n}}_e \; Y^*_{\ell m}(\hat{\mathbf{n}}_e) \; Y^*_{\ell' m'}(\hat{\mathbf{n}}_e) \; Y_{1m''}(\hat{\mathbf{n}}_e) = \sqrt{\frac{(2\ell+1)(2\ell'+1)}{12\pi}} \; \mathrm{C}^{\ell\ell' 1}_{000} \; \mathrm{C}^{\ell\ell' 1}_{mm' m''}. \tag{5.28}$$

These first coefficients $\mathrm{C}^{\ell\ell' 1}_{000}$ are only nonzero for $\ell' = \ell \pm 1$. The second coefficients $\mathrm{C}^{\ell \; \ell\pm1 \; 1}_{mm' m''}$ then require that $m' = m'' - m$ for $m'' = -1, 0, 1$. Therefore, the sums over ℓ' and m' in (5.27) will select six non-zero terms in which (ℓ', m') take the values $(\ell+1, 1-m)$, $(\ell+1, -m)$, $(\ell+$

$1, -1-m)$, $(\ell-1, 1-m)$, $(\ell-1, -m)$, $(\ell-1, -1-m)$. Further, spherical harmonic identities and spherical Bessel recursion relations can then be used to simplify these six terms into just two terms proportional to $Y^*_{\ell m}(\hat{\mathbf{k}})$. Equation (5.27) reduces to

$$\frac{4\pi}{2\ell+1} Y^*_{\ell m}(\hat{\mathbf{k}}) \left[i^{\ell-1} \ \ell \ j_{\ell-1}(k\chi_e) + i^{\ell+1}(\ell+1) j_{\ell+1}(k\chi_e) \right]. \tag{5.29}$$

Plugging this in for the second integral in equation (5.25) leads to the expression

$$a^v_{\ell m}(\chi_e) = \int \frac{d^3k}{(2\pi)^3} \ \Delta^v_\ell(k, \chi_e) \tilde{\Psi}_i(\mathbf{k}) Y^*_{\ell m}(\hat{\mathbf{k}}), \tag{5.30}$$

where we have defined the transfer function $\Delta^v_\ell(k, \chi_e)$ as

$$\Delta^v_\ell(k, \chi_e) \equiv \frac{4\pi \ i^\ell}{2\ell+1} \mathcal{K}^v(k, \chi_e) \left[\ell \ j_{\ell-1}(k\chi_e) - (\ell+1) j_{\ell+1}(k\chi_e) \right] T(k). \tag{5.31}$$

The asymptotic behavior of the transfer function as $k \to \infty$ and $k \to 0$ is given by:

$$\lim_{k\to\infty} \Delta^v_\ell(k, \chi_e) = -\ 4\pi i^\ell \frac{D_v(\chi_e)}{\chi_e} T(k) \cos\left[k\chi_e - \ell\pi/2\right], \tag{5.32}$$

$$\lim_{k\to 0} \Delta^v_\ell(k, \chi_e) = \frac{4\pi \ i^\ell}{2\ell+1} \left[\frac{\ell\sqrt{\pi}(k\chi_e)^{\ell-1}}{2^\ell \Gamma[\frac{1}{2}+\ell]} \right] c_3(\chi_e) k^3, \tag{5.33}$$

where in the small-k limit we have used the fact that $T(0) = 1$ and written the coefficient of the leading order (cubic) term in the Taylor series expansion of $\mathcal{K}^v(k, \chi_e)$ as $c_3(\chi_e)$; in the large-k limit, $T(k) \propto k^{-2}$.

5.3 Simulations

In this section we describe a suite of simulations used to explore the large-scale kSZ effect and provide a concrete example of the relation between the primordial gravitational potential and the effective velocity. This will be used to compute the kSZ signal-to-noise in section 5.6.2. We create three dimensional realizations of the primordial gravitational potential $\Psi(\mathbf{x})$ consistent with ΛCDM (eq. (1.71)) using the method described in Ref. [238] and reviewed in Appendix F. The box size used in each case was $L = 7H_0^{-1} \simeq 31.3$ Gpc. One hundred realizations were created at a resolution of 128^3, covering scales down to $k_{\rm max} \simeq 57.4\ H_0$ ($\lambda_{\rm min} \simeq 484$ Mpc). An example realization is shown in figure 5.3.

With a set of realizations in hand, we then place a hypothetical observer at the center of the box and generate $v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e)$ at 50 equally spaced values of χ_e at a Healpix resolution [239] of $N_{\rm side} = 32$ ($12,288$ equal area pixels of approximately 3.36 square degrees each). This is done as follows. First, we write the effective velocity as

$$v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) = i\hat{\mathbf{n}}_e \cdot \mathbf{V}(\hat{\mathbf{n}}_e, \chi_e), \tag{5.34}$$

where

$$\mathbf{V}(\hat{\mathbf{n}}_e, \chi_e) \equiv \int \frac{d^3k}{(2\pi)^3} \left[T(k)\tilde{\Psi}_i(\mathbf{k})\ \mathcal{K}^v(k, \chi_e)\ \frac{\mathbf{k}}{k} \right]\ e^{i\chi_e \mathbf{k}\cdot\hat{\mathbf{n}}_e}. \tag{5.35}$$

The three components of $\mathbf{V}(\hat{\mathbf{n}}_e, \chi_e)$ can be straightforwardly evaluated for each realization, at each redshift, using a fast Fourier transform (FFT) algorithm. Plugging back into eq. (5.34) and choosing $\hat{\mathbf{n}}_e \cdot \hat{\mathbf{z}} = \cos\theta$, $\hat{\mathbf{n}}_e \cdot \hat{\mathbf{x}} = \sin\theta\cos\phi$, and $\hat{\mathbf{n}}_e \cdot \hat{\mathbf{y}} = \sin\theta\sin\phi$, we interpolate the resulting $v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e)$ at each χ_e onto the Healpix grid. We then take advantage of the Healpix fast spherical harmonic transform functionality to obtain $a^v_{\ell m}(\chi_e)$ at each redshift in each realization.

In the right panel of figure 5.3, we show $v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e)$ at a variety of redshifts in a single realization. In the top row, we choose two nearby redshifts, where it can be seen by eye that there is a good deal of correlation between the two maps. This is because the same large-scale potential field is responsible for the effective velocities at nearby redshifts. In the bottom row, we choose fairly distant redshifts, where the correlation between the two maps is largely absent. Note also the increasing structure with redshift. This is partially due to the limited resolution of the simulation in this figure, but more physically, there is a real effect due to the redshift transfer function. Based on the smallest structures in the resolution probed in our simulation, 64 radial samples would capture all radial structures. Empirically, for all but the smallest angular scale structures, the coherence length between redshift slices is sufficiently long to justify our choice of 50 values of χ_e.

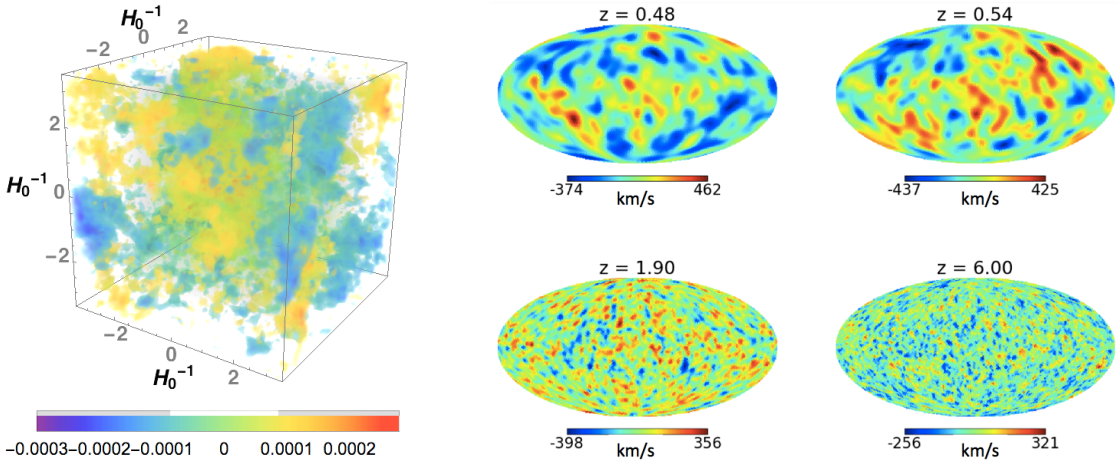

Figure 5.3: By generating random realizations of $\tilde{\Psi}_i(\mathbf{k})$, drawn from a Gaussian distribution with power $P_\Psi(k)$ (parameters given in eq. (1.71)), we can construct realizations for $v_{\rm eff}$. The left panel shows an example realization for $\Psi_i(\mathbf{r})$, and the resulting maps of $v_{\rm eff}$ at various redshifts, using a resolution of $k_{\rm max} \sim 57\ H_0$ ($\lambda_{\rm min} \sim 484$ Mpc). Notice that correlations are evident between the top two $v_{\rm eff}$ maps at close redshift, but not apparent for widely separated redshifts. Also note the increase in structure at higher redshift.

5.4 kSZ Tomography

There is a large amount of information lost in performing the line-of-sight integral in eq. (5.1) for the global kSZ signal. One can in principle do far better by cross correlating the kSZ temperature anisotropies with tracers of the electron density field of known redshift. This is evident in the first detections of the kSZ effect, which were made by isolating the component of the temperature anisotropies associated with the pairwise motion of clusters, whose hot interiors harbor a large density of free electrons. In what follows, we assume the most optimistic scenario possible, in which we have perfect knowledge of the electron density field obtained through the measurement of a completely unbiased tracer. We further assume a purely Gaussian primordial power spectrum, consistent with the current constraints from Planck [8].

To tease out the redshift dependence of the large scale kSZ effect, we introduce a window function $W(\chi_e, \bar{\chi}_e)$ that gives the electron density in a set of redshift bins centered on $\chi_e = \bar{\chi}_e$

$$\delta(\hat{\mathbf{n}}_e, \bar{\chi}_e) = \int d\chi_e W(\chi_e, \bar{\chi}_e)\delta(\hat{\mathbf{n}}_e, \chi_e). \tag{5.36}$$

In the following, we use a top-hat window function normalized to unity: $\int_0^{\chi_\infty} d\chi W(\chi, \bar{\chi}) = 1$. We will consider scenarios with 6, 12, and 24 redshift bins of equal width, covering the range $0 < z < 6$. The redshift coverage for each bin configuration is shown in Figure 5.4.

Forming the cross correlation between the kSZ contribution to the CMB temperature anisotropies and the windowed electron density field, we obtain

$$\begin{aligned}\left\langle \left.\frac{\Delta T}{T}\right|_{\rm kSZ}(\hat{\mathbf{n}}_e)\ \delta(\hat{\mathbf{n}}'_e, \bar{\chi}_e)\right\rangle =& \sigma_T \int d\chi'_e\ W(\chi'_e, \bar{\chi}_e)\ \int d\chi_e\ a(\chi_e)\ \bar{n}_e(\chi_e) \\ &\times \left\langle (1+\delta(\hat{\mathbf{n}}_e, \chi_e))v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e)\delta(\hat{\mathbf{n}}'_e, \chi'_e)\right\rangle. \end{aligned} \tag{5.37}$$

Now comes a very important step. The correlation function above is defined as an ensemble average. Typically, one is interested in using the measured correlation functions to constrain a statistical model of the ensemble. Here, this is not the case. Instead, *we strive to learn information about our particular realization* using the cross correlation, which is information that in the former scenario would have been an obstruction to learning about the theoretical model of the ensemble (e.g. cosmic variance). As we wish to learn about large-scale inhomogeneities, the ensemble average in eq.(5.37) should only be taken over small scales, leaving large scales as a fixed deterministic field.

To formulate this idea more precisely, we decompose the primordial potential into long and short wavelength fields, which can be defined in Fourier space as:

$$\Psi_i^{\mathcal{L}}(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\mathcal{L}(k)\Psi_i(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}, \quad \Psi_i^{\mathcal{S}}(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\mathcal{S}(k)\Psi_i(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{5.38}$$

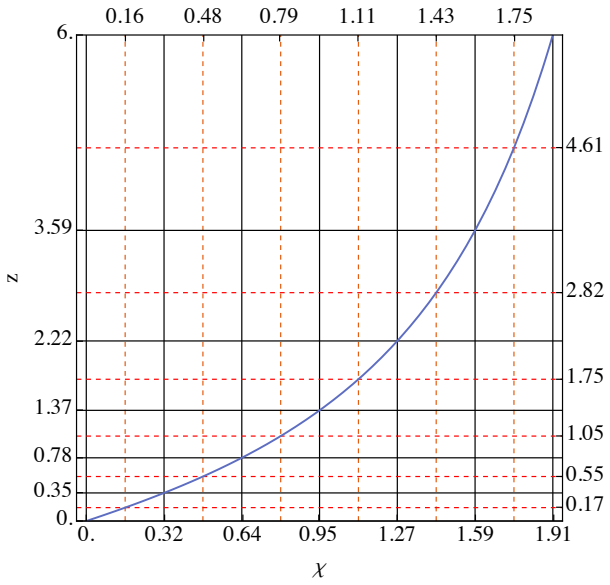

Figure 5.4: Redshift is displayed as a function of comoving distance to illustrate the redshift bin configurations. The solid grid lines show the boundaries that define the redshift bins for $N_{\rm bin} = 6$. The 12 bin configuration is represented by including the dashed grid lines. It is easy to infer the boundary values for $N_{\rm bin} = 24$.

such that $\mathcal{L}(k) + \mathcal{S}(k) = 1$. A suitable choice could be $\mathcal{L}(k) = e^{-k^2/2k_*^2}$, $\mathcal{S}(k) = 1 - e^{-k^2/2k_*^2}$, although we will implicitly be choosing a step function for $\mathcal{L}(k)$ in what follows. Below, we imagine that scales larger than $k_* \gtrsim 10^{-2}$ Mpc^{-1} form the deterministic long field while smaller scales form the stochastic short field. The precise choice of k_* does not affect our results because as we show in section 5.5 and 5.6 small scales ($k \gg 10^{-2}$ Mpc^{-1}) do not contribute to the signal and large scales ($k \ll 10^{-2}$ Mpc^{-1}) do not contribute to the noise.

Because we are working in the linear regime, a long-short split in Ψ_i translates into a long-short split in the effective velocity $v_{\rm eff}$ and the electron density field δ, which we therefore decompose as:

$$v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) = v_{\rm eff}^{\mathcal{L}}(\hat{\mathbf{n}}_e, \chi_e) + v_{\rm eff}^{\mathcal{S}}(\hat{\mathbf{n}}_e, \chi_e), \quad \delta(\hat{\mathbf{n}}_e, \chi_e) = \delta^{\mathcal{L}}(\hat{\mathbf{n}}_e, \chi_e) + \delta^{\mathcal{S}}(\hat{\mathbf{n}}_e, \chi_e). \tag{5.39}$$

Substituting these expansions into the ensemble average in eq. (5.37) and extracting the long wavelength fields in the sense described above, we obtain:

$$\begin{aligned}
\langle (1+\delta)\ v_{\rm eff}\ \delta' \rangle &= \langle (1 + \delta^{\mathcal{L}} + \delta^{\mathcal{S}})\ (v_{\rm eff}^{\mathcal{L}} + v_{\rm eff}^{\mathcal{S}})\ (\delta'^{\mathcal{L}} + \delta'^{\mathcal{S}}) \rangle \\
&= v_{\rm eff}^{\mathcal{L}} \delta'^{\mathcal{L}} + v_{\rm eff}^{\mathcal{L}} \delta^{\mathcal{L}} \delta'^{\mathcal{L}} \\
&\quad + v_{\rm eff}^{\mathcal{L}} \langle \delta'^{\mathcal{S}} \rangle + \delta'^{\mathcal{L}} \langle v_{\rm eff}^{\mathcal{S}} \rangle + \delta^{\mathcal{L}} \delta'^{\mathcal{L}} \langle v_{\rm eff}^{\mathcal{S}} \rangle + v_{\rm eff}^{\mathcal{L}} \delta^{\mathcal{L}} \langle \delta'^{\mathcal{S}} \rangle + v_{\rm eff}^{\mathcal{L}} \delta'^{\mathcal{L}} \langle \delta^{\mathcal{S}} \rangle \\
&\quad + \langle v_{\rm eff}^{\mathcal{S}} \delta^{\mathcal{S}} \delta'^{\mathcal{S}} \rangle \\
&\quad + v_{\rm eff}^{\mathcal{L}} \langle \delta^{\mathcal{S}} \delta'^{\mathcal{S}} \rangle + \delta^{\mathcal{L}} \langle v_{\rm eff}^{\mathcal{S}} \delta'^{\mathcal{S}} \rangle + \delta'^{\mathcal{L}} \langle v_{\rm eff}^{\mathcal{S}} \delta^{\mathcal{S}} \rangle + \langle v_{\rm eff}^{\mathcal{S}} \delta'^{\mathcal{S}} \rangle,
\end{aligned} \tag{5.40}$$

where δ' represents $\delta(\hat{\mathbf{n}}'_e, \chi'_e)$. From our assumption that the short wavelength components are

approximately Gaussian, we set the one-point and three-point correlation functions of short wavelength fields to zero, resulting in the final expression:[1]

$$\begin{aligned}\langle(1+\delta)\ v_{\rm eff}\ \delta'\rangle = & v_{\rm eff}^{\mathcal{L}}\delta'^{\mathcal{L}} + v_{\rm eff}^{\mathcal{L}}\delta^{\mathcal{L}}\delta'^{\mathcal{L}} \\ & + \langle v_{\rm eff}^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle \\ & + v_{\rm eff}^{\mathcal{L}}\langle\delta^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle + \delta^{\mathcal{L}}\langle v_{\rm eff}^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle + \delta'^{\mathcal{L}}\langle v_{\rm eff}^{\mathcal{S}}\delta^{\mathcal{S}}\rangle. \end{aligned} \tag{5.41}$$

The terms on the first line give rise to fluctuations on large angular scales, where the primary CMB dominates. We can therefore eliminate this hopelessly unmeasurable term by filtering the CMB on large angular scales $\ell \lesssim 3000$.[2] The term on the second line gives rise to a statistically isotropic cross power. The terms on the third line give rise to a long wavelength modulation of small-scale power, and will be the focus of what follows. The first of these sources of power asymmetry, $v_{\rm eff}^{\mathcal{L}}\langle\delta^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle$, is far larger than the other two. If we consider a single long ($k^{\mathcal{L}}$) and a single short ($k^{\mathcal{S}}$) wavelength mode, then noting that the Doppler term dominates $v_{\rm eff}^{\mathcal{S}}$ and using $v \propto \delta/k$, we have $\delta^{\mathcal{L}}\langle v_{\rm eff}^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle \sim (k^{\mathcal{L}}/k^{\mathcal{S}})v_{\rm eff}^{\mathcal{L}}\langle\delta^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle \ll v_{\rm eff}^{\mathcal{L}}\langle\delta^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle$. We can therefore approximate

$$\langle(1+\delta)\ v_{\rm eff}\ \delta'\rangle \simeq \langle v_{\rm eff}^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle + v_{\rm eff}^{\mathcal{L}}\langle\delta^{\mathcal{S}}\delta'^{\mathcal{S}}\rangle, \tag{5.42}$$

illustrating that there is a statistically isotropic contribution in the first term that depends only on small scales, and an anisotropic contribution in the second term that depends on large scales. The desired signal is captured in the anisotropic power asymmetry, whereas the small scale isotropic component contributes to the noise which is computed next in section 5.5. Focusing here on the signal, substituting eq. (5.42) into eq. (5.37), and suppressing the $\mathcal{S}$ and $\mathcal{L}$ superscripts, we obtain

$$\begin{aligned}\left\langle \left.\frac{\Delta T}{T}\right|_{\rm kSZ}(\hat{\mathbf{n}}_e)\delta(\hat{\mathbf{n}}'_e,\bar{\chi}_e)\right\rangle = & \sigma_T\int d\chi_e\ a(\chi_e)\ \bar{n}_e(\chi_e)\ v_{\rm eff}(\hat{\mathbf{n}}_e,\chi_e)\int d\chi'_e\ W(\chi'_e,\bar{\chi}_e) \\ & \times\langle\delta(\hat{\mathbf{n}}_e,\chi_e))\ \delta(\hat{\mathbf{n}}'_e,\chi'_e)\rangle + \text{isotropic}. \end{aligned} \tag{5.43}$$

Assuming the electron distribution traces the dark matter, the electron density correlation function is given by

$$\langle\delta(\hat{\mathbf{n}}_e,\chi_e))\ \delta(\hat{\mathbf{n}}'_e,\chi'_e)\rangle = \sum_\ell \frac{2\ell+1}{4\pi}C_\ell^{\delta\delta}(\chi_e,\chi'_e)\mathcal{P}_\ell(\hat{\mathbf{n}}_e\cdot\hat{\mathbf{n}}'_e), \tag{5.44}$$

$$\text{with}\quad C_\ell^{\delta\delta}(\chi_e,\chi'_e) = \int\frac{dk\ k^2}{(2\pi)^3}\ 4\pi\ j_\ell(k\chi_e)\sqrt{P_\delta(k,\chi_e)}\ 4\pi\ j_\ell(k\chi'_e)\sqrt{P_\delta(k,\chi'_e)}, \tag{5.45}$$

where $P_\delta(k,\chi_e)$ is the non-linear matter power spectrum [240], defined in equation (1.75), and computed using the Cosmicpy package.

[1] In any case, significant non-Gaussianity on small scales will not directly mimic the signal we are ultimately after, which is a *long wavelength* modulation of short wavelength power.

[2] The kSZ effect dominates the CMB temperature anisotropies on scales $\ell \gtrsim 3000$, as shown in figure 5.6.

We now arrive at the main conclusion of this section. By isolating the statistically anisotropic term in eq. (5.43), it is possible to measure $v_{\rm eff}$. Because $v_{\rm eff}$ is related to the primordial potential Ψ_i in linear theory by eq. (5.15), this potentially opens a new observational window on large scale inhomogeneities in our Universe. Given a specific model for Ψ_i, it is possible to design an optimal filter to extract the power asymmetry described above [101, 103, 113]. However, in the present context of random Gaussian fields, we quantify the power asymmetry by decomposing into power multipoles [241] that capture the anisotropic term in (5.42),

$$b_{LM}(\bar{\chi}_e) = \int d^2\hat{\mathbf{n}}_e \; Y^*_{LM}(\hat{\mathbf{n}}_e) \left\langle \left.\frac{\Delta T}{T}\right|_{\rm kSZ}(\hat{\mathbf{n}}_e) \; \delta(\hat{\mathbf{n}}_e, \bar{\chi}_e)\right\rangle. \tag{5.46}$$

Let us note that (5.46) is a complete characterization of the signal, as there is no extra information contained in the correlation with $\hat{\mathbf{n}}_e \neq \hat{\mathbf{n}}'_e$. This is clear from (5.43) in the limit of gaussian fields. The next step is to expand the long-wavelength effective velocity into multipoles (see section 5.2.2) using $v_{\rm eff}(\hat{\mathbf{n}}_e, \chi_e) = \sum_{\ell,m} a^v_{\ell m}(\chi_e) Y_{\ell m}(\hat{\mathbf{n}}_e)$ where $a^v_{\ell m}$ is given by (5.30). This makes the angular integral in (5.46) easy to compute as $\int d^2\hat{\mathbf{n}}_e Y^*_{LM}(\hat{\mathbf{n}}_e) Y_{\ell m}(\hat{\mathbf{n}}_e) = \delta_{\ell L}\delta_{mM}$, resulting in the expression,

$$\begin{aligned} b_{LM}(\bar{\chi}_e) = \sum_{\ell=\ell_{\rm min}}^{\ell_{\rm max}} \frac{2\ell+1}{4\pi} \int dk \frac{2k^2}{\pi} \int d\chi_e \; \sigma_T \; a(\chi_e) \; \bar{n}_e(\chi_e) \; a^v_{LM}(\chi_e) \sqrt{P_\delta(k,\chi_e)} \; j_\ell(k\chi_e) \\ \times \int d\chi'_e \; W(\chi'_e, \bar{\chi}_e) \sqrt{P_\delta(k,\chi'_e)} \; j_\ell(k\chi'_e). \end{aligned} \tag{5.47}$$

We therefore see that the power multipoles in each bin are proportional to a weighted integral of the corresponding multipole of the projected effective velocity field. The lower and upper bounds on the summation, $(\ell_{\rm min},\ \ell_{\rm max})$, reflect the filtering and resolution scales that might be achievable. By default, we will use $(\ell_{\rm min} = 3000,\ \ell_{\rm max} = \infty)$ unless otherwise stated. The effects of varying these bounds will be discussed in section 5.8.

5.5 Cosmic Variance Limited Noise

In this section we estimate the cosmic variance limited noise that we expect for the power multipoles computed in the previous section. However, above, our focus was the statistically anisotropic contribution to the cross correlation, but here we are interested in the statistically isotropic contribution to temperature anisotropies, which depends predominantly on small scales.

On small angular scales, the late-time kSZ effect is the dominant source of temperature anisotropies in the CMB (see the right panel of figure 5.6 below). Under the assumption of approximate Gaussianity, and assuming a perfectly unbiased measurement of the electron

density field, the primary source of noise is therefore an "accidental" power asymmetry in the cross correlation. We can estimate this through the variance in the power multipoles, which is computed as the coincident limit of the four point function between two powers of $\frac{\Delta T}{T}$ and two powers of δ. Specifically, we must compute:

$$\begin{aligned}\langle \tilde{b}^*_{LM}(\bar{\chi}_e)\tilde{b}_{LM}(\bar{\chi}_e)\rangle = &\int d^2\hat{\mathbf{n}}_e d^2\hat{\mathbf{n}}'_e\ Y^*_{LM}(\hat{\mathbf{n}}_e)\ Y_{LM}(\hat{\mathbf{n}}'_e)\\ &\times\Big\langle \frac{\Delta T}{T}(\hat{\mathbf{n}}_e)\ \delta(\hat{\mathbf{n}}_e,\bar{\chi}_e)\ \frac{\Delta T}{T}(\hat{\mathbf{n}}'_e)\ \delta(\hat{\mathbf{n}}'_e,\bar{\chi}_e)\Big\rangle,\end{aligned} \tag{5.48}$$

where we are using tildes on the b_{LM}'s to indicate that these are not simply the same power multipoles as in (5.47). Instead, the variance here captures the chance power asymmetry that is present in the statistically isotropic contribution to $\frac{\Delta T}{T}$, which is sensitive only to small scales where the "conventional" kSZ effect dominates. Since the primary CMB also causes statistical fluctuations in the anisotropic kSZ measurement, we consider both kSZ and primary CMB contributions to $\frac{\Delta T}{T}$:

$$\frac{\Delta T}{T} = \frac{\Delta T}{T}\bigg|_{\rm kSZ} + \frac{\Delta T}{T}\bigg|_{\rm p} \tag{5.49}$$

The 4-point function has contributions from all possible 2-point functions and an irreducible/connected piece (which we assume to be negligibly small). Noting that the cross correlation of the primary CMB and the density field is zero, we obtain,

$$\begin{aligned}&\Big\langle \frac{\Delta T}{T}(\hat{\mathbf{n}}_e)\ \delta(\hat{\mathbf{n}}_e,\bar{\chi}_e)\ \frac{\Delta T}{T}(\hat{\mathbf{n}}'_e)\ \delta(\hat{\mathbf{n}}'_e,\bar{\chi}_e)\Big\rangle\\ =&\Big\langle \frac{\Delta T}{T}(\hat{\mathbf{n}}_e)\ \frac{\Delta T}{T}(\hat{\mathbf{n}}'_e)\Big\rangle\Big\langle \delta(\hat{\mathbf{n}}_e,\bar{\chi}_e)\ \delta(\hat{\mathbf{n}}'_e,\bar{\chi}_e)\Big\rangle\\ &+\Big\langle \frac{\Delta T}{T}\bigg|_{\rm kSZ}(\hat{\mathbf{n}}_e)\ \delta(\hat{\mathbf{n}}'_e,\bar{\chi}_e)\Big\rangle\Big\langle \frac{\Delta T}{T}\bigg|_{\rm kSZ}(\hat{\mathbf{n}}'_e)\ \delta(\hat{\mathbf{n}}_e,\bar{\chi}_e)\Big\rangle.\end{aligned} \tag{5.50}$$

Assuming the electron density field traces dark matter, the electron density autocorrelation function is given by

$$\Big\langle \delta(\hat{\mathbf{n}}_e,\bar{\chi}_e)\ \delta(\hat{\mathbf{n}}'_e,\bar{\chi}_e)\Big\rangle = \sum_\ell \frac{2\ell+1}{4\pi}\ C^{\delta\delta}_\ell\ \mathcal{P}_\ell(\hat{\mathbf{n}}_e\cdot\hat{\mathbf{n}}'_e), \tag{5.51}$$

$$\begin{aligned}\text{with}\quad C^{\delta\delta}_\ell &= \int d\chi_e\ W(\chi_e,\bar{\chi}_e)\int d\chi'_e\ W(\chi'_e,\bar{\chi}_e)\ C^{\delta\delta}_\ell(\chi_e,\chi'_e)\\ &= \int dk\frac{2k^2}{\pi}\int d\chi_e\sqrt{P_\delta(k,\chi_e)}W(\chi_e,\bar{\chi}_e)j_\ell(k\chi_e)\\ &\quad\times\int d\chi'_e\sqrt{P_\delta(k,\chi'_e)}W(\chi'_e,\bar{\chi}_e)j_\ell(k\chi'_e)\\ &\simeq \int\frac{dk}{\ell+1/2}W^2\left(\frac{\ell+1/2}{k},\bar{\chi}_e\right)P_\delta\left(k,\frac{\ell+1/2}{k}\right),\end{aligned} \tag{5.52}$$

where we used the expression for $C^{\delta\delta}_\ell(\chi_e,\chi'_e)$ from (5.45), and the Limber approximation [242] in the last line. The quantity $C^{\delta\delta}_\ell$ is shown in figure 5.5.

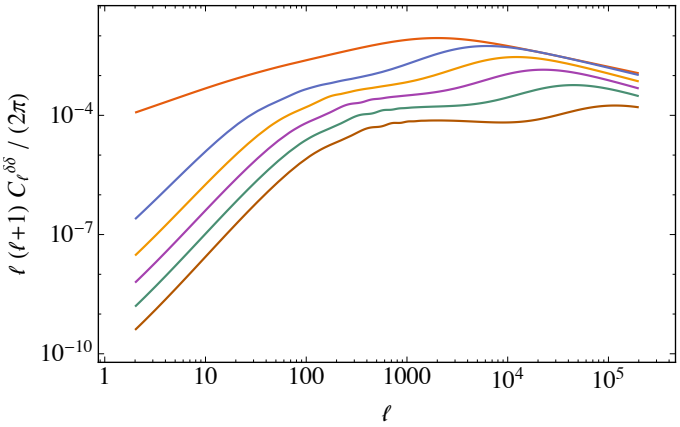

Figure 5.5: The non-linear angular matter power spectrum computed in six different redshift bins described by the solid lines in figure 5.4 from today (top) to $z = 6$ (bottom).

The temperature autocorrelation function has contributions from the primary CMB and the conventional kSZ effect. The latter dominates past $\ell \sim 3000$ and has the following form:

$$\left\langle \frac{\Delta T}{T}\bigg|_{\rm kSZ}(\hat{\mathbf{n}}_e)\ \frac{\Delta T}{T}\bigg|_{\rm kSZ}(\hat{\mathbf{n}}'_e)\right\rangle = \int d\chi_e \sigma_T a(\chi_e)\bar{n}_e(\chi_e)\int d\chi'_e \sigma_T a(\chi'_e)\bar{n}_e(\chi'_e) \times \langle q(\chi_e,\hat{\mathbf{n}}_e)q(\chi'_e,\hat{\mathbf{n}}'_e)\rangle \tag{5.53}$$

$$= \sum_\ell \frac{2\ell+1}{4\pi}\, C_\ell^{TT,\rm kSZ}\, \mathcal{P}_\ell(\hat{\mathbf{n}}_e\cdot\hat{\mathbf{n}}'_e), \tag{5.54}$$

where $q \equiv \mathbf{q}\cdot\hat{\mathbf{n}}_e \equiv v_{\rm eff}(1+\delta)$ is the momentum field of free electrons. To obtain an expression for the temperature power spectrum, $C_\ell^{TT,\rm kSZ}$, the key quantity to compute is $\langle q(\chi_e,\hat{\mathbf{n}}_e)q(\chi'_e,\hat{\mathbf{n}}'_e)\rangle$, which is the fourth moment of two δ's and two v's. Schematically, $\langle qq\rangle = \langle vv\rangle\langle\delta\delta\rangle + 2\langle v\delta\rangle^2 + \langle v\delta v\delta\rangle_c$, where the subscript c denotes the irreducible connected term. The momentum power spectrum, denoted by P_q, is typically computed by decomposing the Fourier transform, $\tilde{\mathbf{q}}(\mathbf{k})$, into components parallel to $\hat{\mathbf{k}}$, $\tilde{\mathbf{q}}_\parallel = \hat{\mathbf{k}}(\tilde{\mathbf{q}}\cdot\hat{\mathbf{k}})$, and perpendicular to $\hat{\mathbf{k}}$, $\tilde{\mathbf{q}}_\perp = \tilde{\mathbf{q}} - \hat{\mathbf{k}}(\tilde{\mathbf{q}}\cdot\hat{\mathbf{k}})$ [243]. The longitudinal momentum component does not contribute significantly to $C_\ell^{TT,\rm kSZ}$ due to cancellations of positive and negative contributions in the line-of-sight integration. For instance, as shown by Park et al. [244], the longitudinal contribution to $C_\ell^{TT,\rm kSZ}$ is more than four orders of magnitude below the transverse contribution for $\ell > 3000$, so it suffices to only consider $P_{q_\perp}$ in this calculation.

To compute $P_{q_\perp}$, we use the "standard kSZ model" [244, 245], which incorporates the fully non-linear power spectrum for the density field, $P_\delta^{\rm nl}$, but approximates the velocity power spectrum by linear theory $P_v = \frac{\dot{a}^2 f^2}{k^2} P_\delta^{\rm lin}$ where $f = \left(1 + \frac{a}{D_\Psi}\frac{dD_\Psi}{da}\right)^2$. The resulting expression

is

$$P_{q_\perp}(k,\chi) = \dot{a}^2 f^2 \int_{-1}^{1} d\mu \int \frac{dk'}{(2\pi)^2}\; P_\delta^{\rm lin}(k',\chi)\; P_\delta^{\rm nl}(\sqrt{k^2-2kk'\mu+k'^2},\chi) \times \frac{(k^2-2kk'\mu)(1-\mu^2)}{k^2-2kk'\mu+k'^2}. \tag{5.55}$$

This model neglects the velocity-density cross correlation because the geometrical factor attached to this term decreases rapidly at large k. Since the kSZ contribution to C_ℓ^{TT} that we are interested in is sensitive only to high k scales, this approximation is valid. Note that this also implies our precise choice of k_* in section 5.4 is irrelevant for the noise calculation. We are also neglecting the non-Gaussian contribution from the connected 4-point function (unimportant on all scales according to [243], but could account for up to 10% of the power spectrum according to [244]). In terms of $P_{q_\perp}$ the temperature power spectrum can be expressed as

$$\begin{aligned} C_\ell^{TT,\rm kSZ} &= \int dk \frac{2k^2}{\pi} \int d\chi_e \sigma_T a(\chi_e)\bar{n}_e(\chi_e)\sqrt{P_{q_\perp}(k,\chi_e)}\, j_\ell(k\chi_e) \\ &\quad \times \int d\chi'_e \sigma_T a(\chi'_e)\bar{n}_e(\chi'_e)\sqrt{P_{q_\perp}(k,\chi'_e)}\, j_\ell(k\chi'_e) \\ &\simeq \int \frac{dk}{\ell+1/2} \left[\sigma_T a(\chi)\bar{n}_e(\chi)\right]^2 P_{q_\perp}(k,\chi)\Big|_{\chi\to(\ell+1/2)/k}, \end{aligned} \tag{5.56}$$

using the Limber approximation [242] again. The functions $P_{q_\perp}(k,\chi)$ and $C_\ell^{TT,\rm kSZ}$ are shown in figure 5.6. The total temperature power spectrum is a sum of the primary and kSZ contributions, denoted from now on by

$$C_\ell^{TT} = C_\ell^{TT,\rm p} + C_\ell^{TT,\rm kSZ}. \tag{5.57}$$

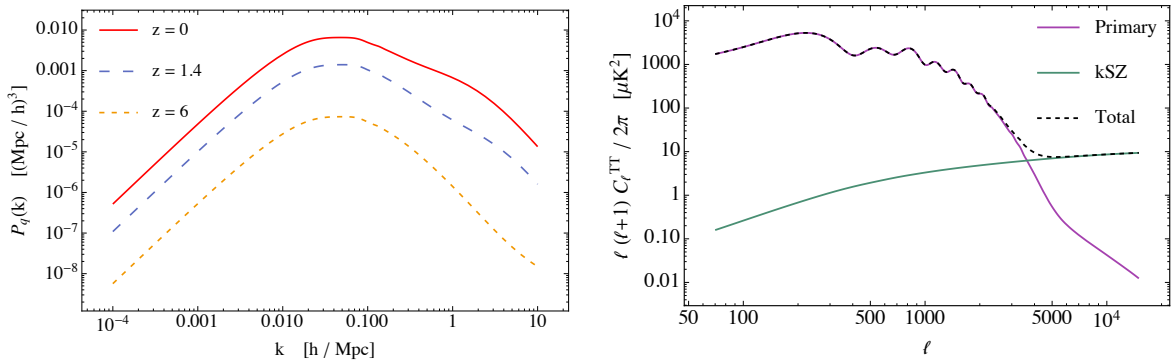

Figure 5.6: Left: the power spectrum of the transverse momentum field (5.55). Right: the contributions to the temperature-temperature power spectrum from the primary CMB and the non-linear kSZ effect (5.56).

The second term in eq. (5.50) requires us to compute the cross correlation of the temperature

with density. This can be done as follows

$$\begin{aligned}\left\langle \frac{\Delta T}{T}\Big|_{\rm kSZ}(\hat{\mathbf{n}}_e)\delta(\hat{\mathbf{n}}'_e,\bar{\chi}_e)\right\rangle &= -\int d\chi_e \sigma_T a(\chi_e)\bar{n}_e(\chi_e)\langle v_{\rm eff}(\hat{\mathbf{n}}_e,\chi_e)(1+\delta(\hat{\mathbf{n}}_e,\chi_e))\delta(\hat{\mathbf{n}}'_e,\bar{\chi}_e)\rangle \\ &= -\int d\chi_e \sigma_T a(\chi_e)\bar{n}_e(\chi_e)\int d\chi'_e W(\chi'_e,\bar{\chi}_e)\ \langle v_{\rm eff}(\hat{\mathbf{n}}_e,\chi_e)\delta(\hat{\mathbf{n}}'_e,\chi'_e)\rangle \\ &= \sum_\ell \frac{2\ell+1}{4\pi}\ C_\ell^{T\delta}\ \mathcal{P}_\ell(\hat{\mathbf{n}}_e\cdot\hat{\mathbf{n}}'_e). \end{aligned} \tag{5.58}$$

Assuming approximate Gaussianity, we neglect the three point function $\langle v\delta\delta\rangle$, so only the correlation between $v_{\rm eff}$ and δ remains. Recall that we are computing the contribution from the small scales, whereas the signal calculation above (see eqs. (5.42),(5.43)) is only sensitive to large scales. The power spectrum $C_\ell^{T\delta}$ takes the form

$$\begin{aligned} C_\ell^{T\delta} &= -\int \frac{dk\ k^2}{(2\pi)^3}\int d\chi_e \sigma_T a(\chi_e)\bar{n}_e(\chi_e)\Delta_\ell^v(k,\chi_e)\sqrt{P_\Psi(k)} \\ &\quad\times\int d\chi'_e 4\pi\sqrt{P_\delta(k,\chi'_e)}W(\chi'_e,\bar{\chi}_e)j_\ell(k\chi'_e) \\ &= -\int \frac{dk\ k^2}{(2\pi)^3}\int d\chi_e \sigma_T a(\chi_e)\bar{n}_e(\chi_e)4\pi\mathcal{K}^v(k,\chi_e)\frac{T(k)}{k}\frac{dj_\ell(k\chi_e)}{d\chi_e}\sqrt{P_\Psi(k)} \\ &\quad\times\int d\chi'_e 4\pi\sqrt{P_\delta(k,\chi'_e)}W(\chi'_e,\bar{\chi}_e)j_\ell(k\chi'_e), \end{aligned} \tag{5.59}$$

where we've recalled the expression for $\Delta_\ell^v(k,\chi_e)$ from eq. (5.31), and used the identity,

$$\frac{\ell\ j_{\ell-1}(k\chi)-(\ell+1)j_{\ell+1}(k\chi)}{2\ell+1} = \frac{1}{k}\frac{dj_\ell(k\chi)}{d\chi}. \tag{5.60}$$

Integrating by parts and using the Limber approximation [242] results in the expression

$$\begin{aligned} C_\ell^{T\delta} &= \int dk\frac{2k^2}{\pi}\int d\chi_e\sigma_T\frac{d}{d\chi_e}[a(\chi_e)\bar{n}_e(\chi_e)\mathcal{K}^v(k,\chi_e)]\frac{T(k)}{k}\sqrt{P_\Psi(k)}j_\ell(k\chi_e) \\ &\quad\times\int d\chi'_e\sqrt{P_\delta(k,\chi'_e)}W(\chi'_e,\bar{\chi}_e)j_\ell(k\chi'_e) \\ &\simeq \int\frac{dk}{\ell+1/2}\ \sigma_T\ \frac{d}{d\chi}[a(\chi)\bar{n}_e(\chi)\mathcal{K}^v(k,\chi)]\ \frac{T(k)}{k}\ \sqrt{P_\Psi(k)P_\delta(k,\chi)}\ W(\chi,\bar{\chi}_e)\Big|_{\chi\to(\ell+1/2)/k}. \end{aligned} \tag{5.61}$$

Combining all of these pieces, expanding $\mathcal{P}_\ell(\hat{\mathbf{n}}_e\cdot\hat{\mathbf{n}}'_e)$ in terms of spherical harmonics using the identity (5.26), the variance in the power multipoles can then be written

$$\begin{aligned}\langle\tilde{b}_L^2\rangle &= \sum_M\frac{\langle\tilde{b}^*_{LM}\tilde{b}_{LM}\rangle}{2L+1} \\ &= \sum_{\ell,\ell',m,m',M}\frac{C_\ell^{TT}C_{\ell'}^{\delta\delta}+C_\ell^{T\delta}C_{\ell'}^{T\delta}}{2L+1}\ \Big|\int d^2\hat{\mathbf{n}}_e Y_{LM}(\hat{\mathbf{n}}_e)Y^*_{\ell m}(\hat{\mathbf{n}}_e)Y^*_{\ell' m'}(\hat{\mathbf{n}}_e)\Big|^2 \\ &= \sum_{\ell,\ell'}[C_\ell^{TT}C_{\ell'}^{\delta\delta}+C_\ell^{T\delta}C_{\ell'}^{T\delta}]\ \frac{(2\ell+1)(2\ell'+1)}{4\pi(2L+1)^2}\ \Big|\mathrm{C}^{\ell\ell' L}_{000}\Big|^2\sum_{m,m',M}\Big|\mathrm{C}^{\ell\ell' L}_{mm'M}\Big|^2 \\ &= \sum_{\ell,\ell'}[C_\ell^{TT}C_{\ell'}^{\delta\delta}+C_\ell^{T\delta}C_{\ell'}^{T\delta}]\ \frac{(2\ell+1)(2\ell'+1)}{4\pi(2L+1)}\ \Big|\mathrm{C}^{\ell\ell' L}_{000}\Big|^2, \end{aligned} \tag{5.62}$$

where $\mathrm{C}^{\ell\ell' L}_{mm' M}$ are the Clebsch-Gordan coefficients, and we used the triple product integral identity (see eq. (5.28)) and $\sum_{m,m',M} \left|\mathrm{C}^{\ell\ell' L}_{mm' M}\right|^2 = 2L+1$. Note that the coefficients $\mathrm{C}^{\ell\ell' L}_{000}$ are only nonzero for $|\ell - \ell'| \leq L \leq \ell + \ell'$ and $\ell + \ell' + L$ even. Figure 5.7 shows $(C_\ell^{T\delta})^2$ plotted in three different redshift bins, and compares this to $C_\ell^{TT} C_\ell^{\delta\delta}$. It is clear that the cross term is subdominant by several orders of magnitude for all ℓ, and therefore it is sufficient to approximate $C_\ell^{TT} C_{\ell'}^{\delta\delta} + C_\ell^{T\delta} C_{\ell'}^{T\delta} \sim C_\ell^{TT} C_{\ell'}^{\delta\delta}$.

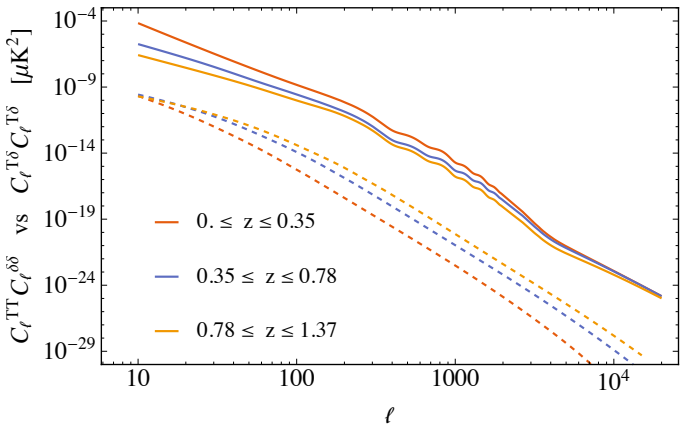

Figure 5.7: The solid curves are $C_\ell^{TT} C_\ell^{\delta\delta}$ and the dashed curves are $C_\ell^{T\delta} C_\ell^{T\delta}$, which represent the two terms in square brackets in eq. (5.62) evaluated at the same ℓ. The redshift ranges indicate sample redshift bins (see figure 5.4).

The final result for the cosmic variance limited noise estimate is

$$\sqrt{\langle \tilde{b}_L(\bar{\chi}_e)^2 \rangle} = \sqrt{\sum_{\ell,\ell'=\ell_{\min}}^{\ell_{\max}} [C_\ell^{TT} C_{\ell'}^{\delta\delta}(\bar{\chi}_e)] \, \frac{(2\ell+1)(2\ell'+1)}{4\pi(2L+1)} \left|\mathrm{C}^{\ell\ell' L}_{000}\right|^2}. \tag{5.63}$$

where we have restored an explicit dependence on the redshift bin $\bar{\chi}_e$. Notice that the result is given by a sum over ℓ, ℓ', and recall that the noise captures the accidental power asymmetry which is sensitive only to high ℓ. In the spirit of our cosmic variance limited estimate, we can imagine that we have sufficiently filtered out the primary CMB and cleaned the foregrounds. We capture this in the calculation by starting the sum in (5.63) at $\ell_{\min} = 3000$. As previously mentioned, we will employ the default values ($\ell_{\min} = 3000$, $\ell_{\max} = \infty$) unless otherwise stated. Results for varying $\ell_{\min}$ and $\ell_{\max}$ will be presented in section 5.8, allowing us to make statements about the detectability with next generation CMB experiments and galaxy surveys.

5.6 Cosmic Variance Limited Signal to Noise

In this section, we assess the signal to noise for the power multipoles eq. (5.47) in the cosmic variance limit using both a theoretical estimate and the simulations described in section 5.3.

In each case, the signal to noise in each bin $\bar{\chi}_e$ is calculated as

$$\frac{S}{N}(\bar{\chi}_e) = \frac{b_{LM}(\bar{\chi}_e)}{\sqrt{\langle \tilde{b}_L(\bar{\chi}_e)^2 \rangle}} \tag{5.64}$$

where $b_{LM}(\bar{\chi}_e)$ is found using eq. (5.47) and $\sqrt{\langle \tilde{b}_L(\bar{\chi}_e)^2 \rangle}$ is given by eq. (5.63).

5.6.1 RMS Estimate

A simple estimate of the signal is obtained by approximating $a^v_{LM}(\chi_e) \sim \sqrt{C^v_L(\chi_e)}$ in eq. (5.47), where $C^v_L(\chi_e)$ is the power spectrum associated with the large-scale velocity, given by

$$C^v_L(\chi_e) = \int_0^{k_{\rm max}} \frac{k^2 dk}{(2\pi)^3} P_\Psi(k) |\Delta^v_L(k, \chi_e)|^2 \tag{5.65}$$

with $\Delta^v_L(k, \chi_e)$ given by eq. (5.31). This necessarily yields an overestimate of the signal, since in reality $a^v_{L0}(\chi_e)$ will vary over the window functions leading to partial cancellation, while here we are assuming that it always takes its (positive definite) RMS value. By comparing with simulations in the following subsection, we show that this approximation gives a good estimate in the limit of thin window functions $W(\chi_e, \bar{\chi}_e)$ (in the context of this work, this is equivalent to the limit of many redshift bins).

Before proceeding, we can assess which scales form the dominant contribution to eq. (5.65). The upper limit of integration in eq. (5.65), $k_{\rm max}$, corresponds to the smallest scale, $\lambda_{\rm min} = 2\pi/k_{\rm max}$, that contributes to the signal. Formally, $k_{\rm max} \to \infty$, but we can adjust the cutoff to include only the long modes discussed above. In figure 5.8 we show C^v_L at $z = 1$ as a function of $k_{\rm max}$. Vertical lines indicate $\lambda_{\rm min} = 10, 10^2, 10^3$ Mpc. Here, we see that for a number of power multipoles, the relevant signal is obtained almost entirely from scales $\sim 10^2 - 10^3$ Mpc.

Putting everything together, applying the Limber approximation [242], and choosing $M = 0$ under the assumption that all other M will statistically be the same, the final expression for the signal becomes

$$b_{L0}(\bar{\chi}_e) \simeq \sum_{\ell=\ell_{\rm min}}^{\ell_{\rm max}} \frac{2\ell+1}{4\pi} \int \frac{d\chi}{\chi^2}\, \sigma_T\, a(\chi)\, \bar{n}_e(\chi)\, \sqrt{C^v_L(\chi)}\, W(\chi, \bar{\chi}_e)\, P_\delta\left(\frac{\ell+1/2}{\chi}, \chi\right). \tag{5.66}$$

We compute the signal eq. (5.66) and noise eq. (5.63) for the $N_{\rm bin} = 6,\ 12,$ and 24 top-hat bin configurations with redshift ranges summarized in figure 5.4. We show the results for the 6-bin configuration in figure 5.9. Each plot has four curves. The solid red curve is the RMS signal with $k_{\rm max} \to \infty$, the dotted orange line is the cosmic variance limited noise, the short dashed blue line is the RMS signal computed for $k_{\rm max} = 278\ H_0$ ($\lambda_{\rm min} = 100$ Mpc), and the long dashed green curve is the signal computed for $k_{\rm max} = 2780\ H_0$ ($\lambda_{\rm min} = 10$ Mpc).

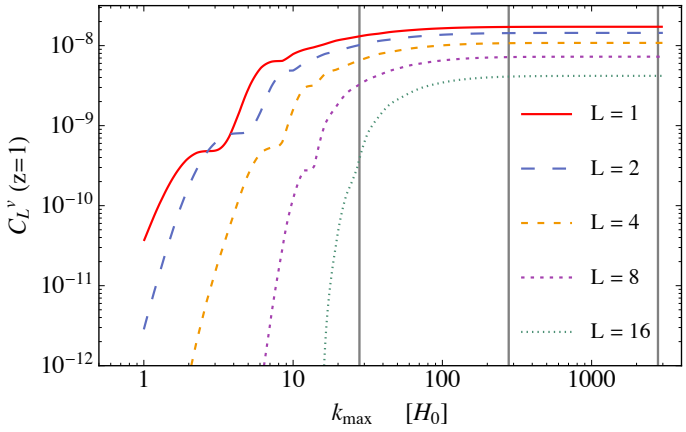

Figure 5.8: The effective velocity contribution to the signal described in eq. (5.65) versus $k_{\max}$ in units of H_0, which refers to the upper bound on the integral. Notice that the contribution mainly comes from large scales (small k) for low L. The vertical lines indicate scales equal to 1 Gpc, 100 Mpc, and 10 Mpc from left to right.

Comparing the three signal curves we see that except in the lowest redshift bin at high L, the signal is composed primarily of long wavelength modes ($\lambda > 100$ Mpc), as expected. The amplitude of the signal varies by roughly an order of magnitude between the lowest and highest redshift bin, and is strongest at low L and low redshift. Both the signal and noise are relatively flat over the plotted range in L. Most importantly though, the signal is 2-3 orders of magnitude larger than the noise!

To compare the result for the 12 and 24 bin configurations, in figure 5.10 we show the signal to noise eq. (5.64) at $L = 1$ and $L = 50$ as a function of comoving distance. Increasing the number of bins by a factor of 2 results in a decrease in the signal-to-noise by a factor of $\sim \sqrt{2}$. This is true for all L. We therefore conclude that a signal can in principle be measured in the cosmic variance limit at high signal to noise for a large number of redshift bins $N_{\rm bins}$ at a variety of scales L.

5.6.2 Comparing with the signal from simulations

The above calculation is an over-estimate of the signal, as it assumes the velocity field is positive definite and given by its RMS value. In particular, it does not account for partial cancellations along the line of sight. To take this into account, we can compute the signal from eq. (5.47) using the effective velocity field computed from the simulations described in section 5.3. Using an interpolating function for $a^v_{\ell m}(\chi_e)$ constructed from the $v_{\rm eff}$ map in each of the 50 redshift bins, we compute eq. (5.47) directly for 100 realizations at a resolution of $k_{\max} \sim 57.4\ H_0$ for the 6 and 12 bin configurations. Below we only present results for b_{L0}; other values of M have identical statistical properties.

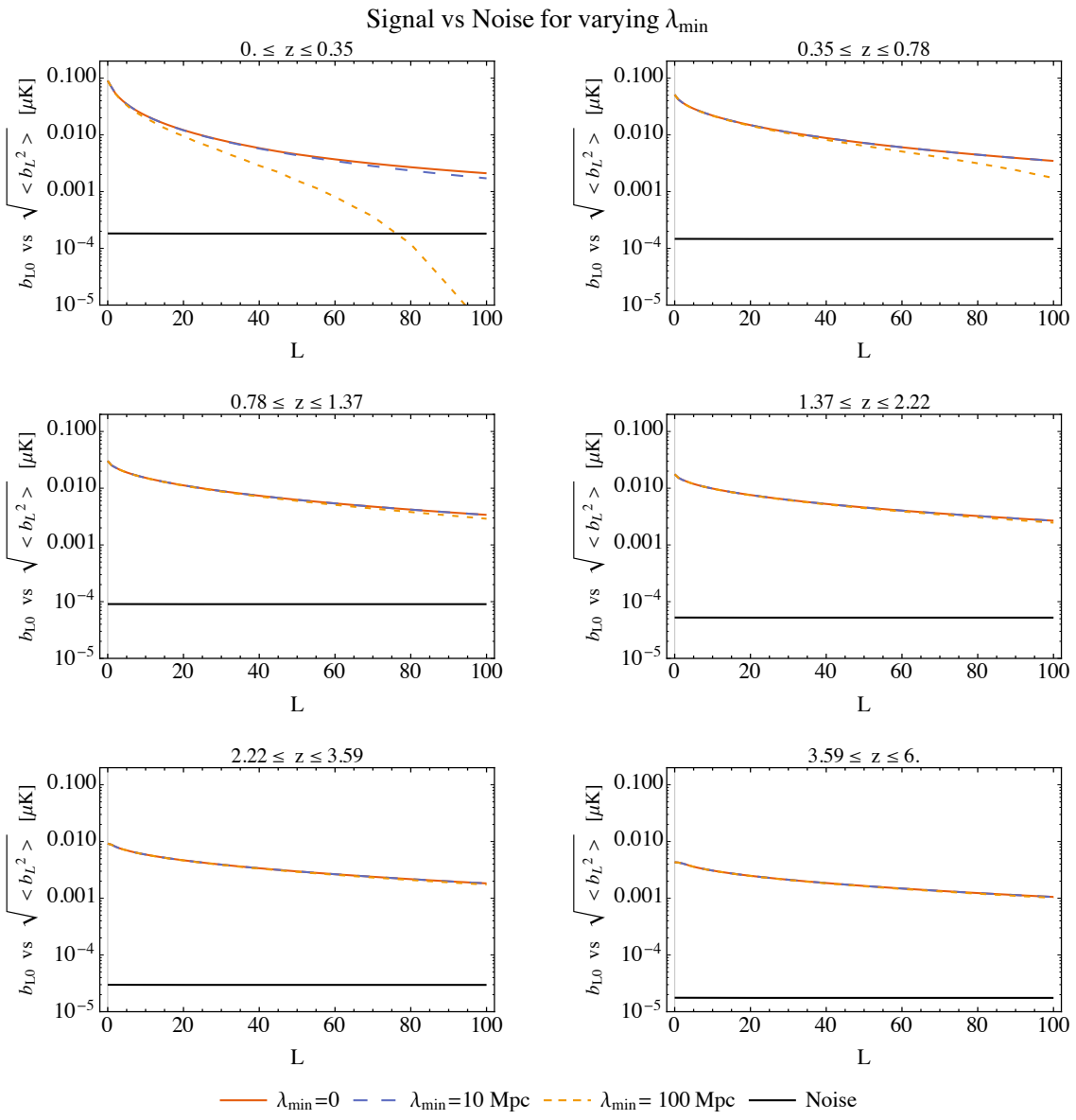

Figure 5.9: The signal (5.66) is shown in six redshift bins, calculated for three different scales: $k_{\max} = 278\ H_0$ ($\lambda_{\min} = 100$ Mpc) [orange, short dash], $k_{\max} = 2780\ H_0$ ($\lambda_{\min} = 10$ Mpc) [blue, long dash], and $k_{\max} = \infty$ ($\lambda_{\min} = 0$) [red, solid]. The solid black line is the noise estimate $\sqrt{\langle \tilde{b}_L^2 \rangle}$ (eq. (5.63)), which falls well below the estimated signal for this configuration. The summation bounds employed for the signal and noise calculations are ($\ell_{\min} = 3000,\ \ell_{\max} = \infty$).

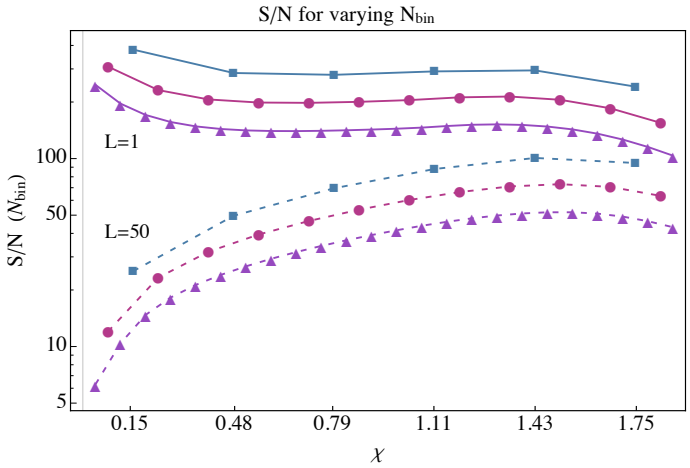

Figure 5.10: The bullet points (square, circle, triangle) represent the signal-to-noise (eq. (5.64)) in each bin for $N_{\rm bin} = (6, 12, 24)$, plotted against χ at the midpoint of the bin. The solid top curves are for $L = 1$ and the bottom dashed curves are for $L = 50$. The signal and noise were computed using eq. (5.66) and (5.63), employing summation bounds ($\ell_{\rm min} = 3000$, $\ell_{\rm max} = \infty$), and with $k_{\rm max} \to \infty$ in eq. (5.65).

Figures 5.11 and 5.12 show the signal to noise computed using simulations, in comparison to the one estimated from theory with eq. (5.65) computed using the integration limit $k_{\rm max} \sim 57.4\ H_0$ corresponding to the simulation resolution. We plot $L < 10$, which is accurately captured for the resolution we consider (see figure 5.8). Notice that the solid curves, showing the predicted signal based on our theory calculation, are always higher than the average signal from the realizations. This is due to a difference in the order of operations. In our estimation in (5.66) we have averaged over $a_{L0}(\chi_e)$ prior to integrating over χ_e, whereas the signal computed from the realizations integrates over χ_e first and then averages. Since $a_{L0}(\chi_e)$ is an oscillating function that takes positive and negative values, there can be cancellation upon integration over χ_e. This cancellation can be minimized by using smaller redshift bins, which results in a smaller range of integration and a lesser chance for cancellation. This can be noticed empirically as the agreement between realizations and theory is better for the 12 bin configuration than the 6 bin. In summary, the realizations approach the theory estimation for more bins as a result of having less variation in a_{LM} over the bin.

5.7 Mode counting

With an understanding of this signal in hand, we now want to estimate how many modes one could conceivably measure in the cosmic variance limit. Let's consider scales $\lambda > 100$ Mpc, of order the BAO scale. This corresponds to $k_{\rm max} = 278\ H_0$ and $L_{\rm max} \sim \pi/\theta \sim \pi\chi_{\rm dec}/\lambda_{\rm min} = k_{\rm max}\chi_{\rm dec}/2 \sim 437$. On this scale and larger, the primary CMB contains $\sum_L^{437}(2L+1) = 191843$

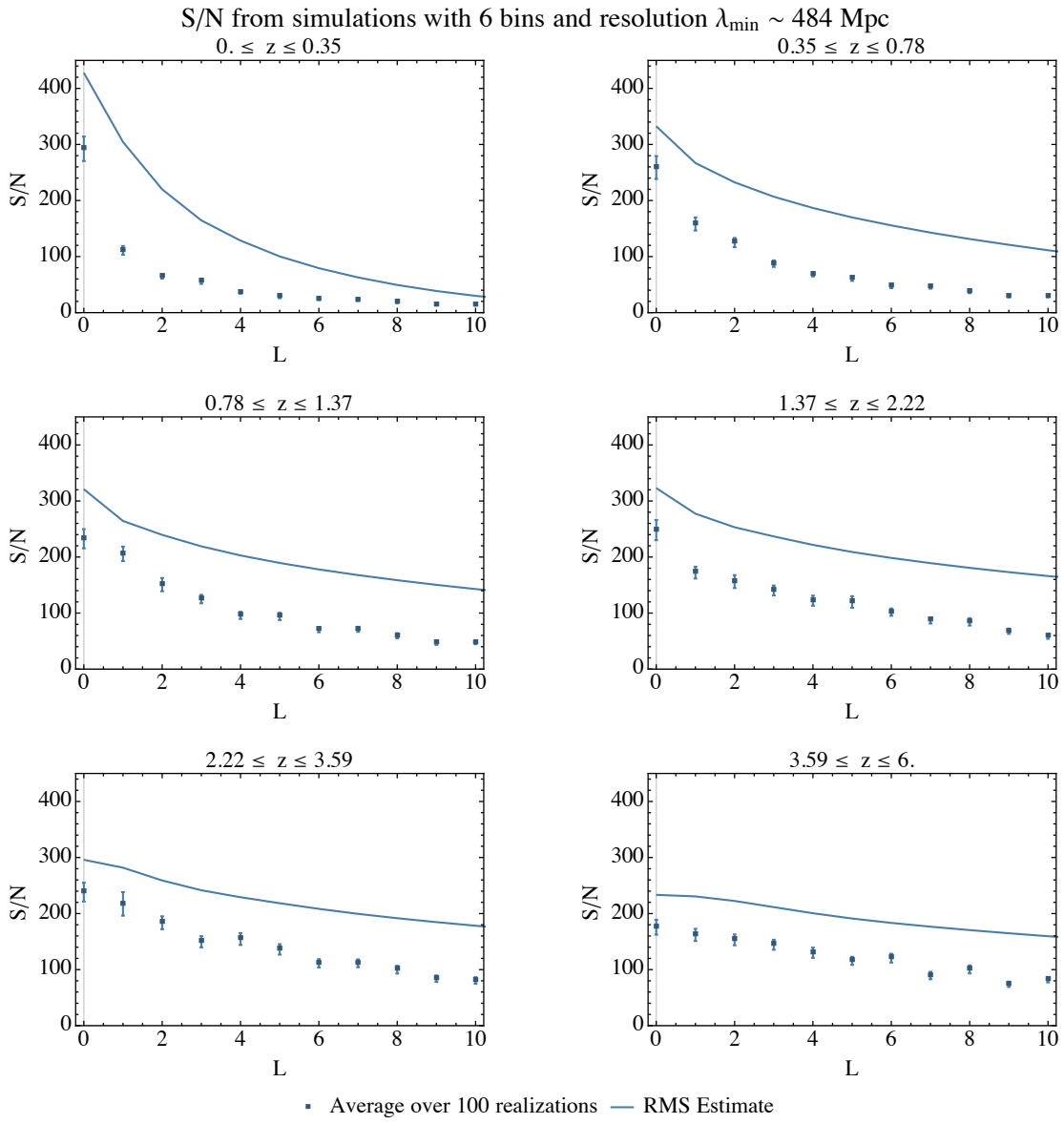

Figure 5.11: The signal-to-noise (eq. (5.64)) computed using simulations, in comparison to the RMS estimate (solid curves), in 6 redshift bins (see figure 5.4). The points represent the standard deviation of the 100 realizations, and the error bars denote the standard error of the standard deviation. These simulations have a resolution of $\lambda_{\min} \sim 484$ Mpc and utilized the default summation bounds ($\ell_{\min} = 3000$, $\ell_{\max} = \infty$) for the signal and noise calculations.

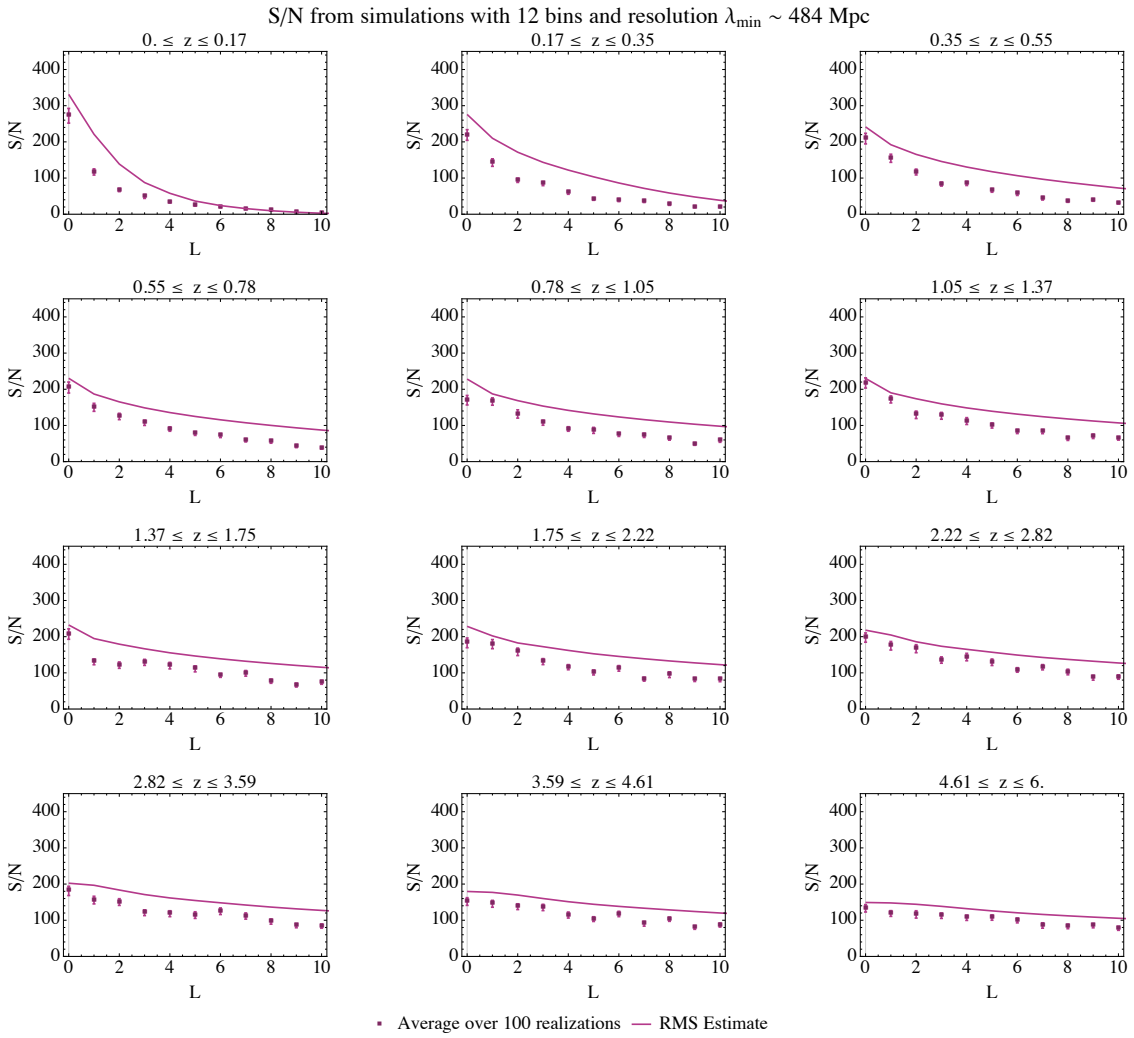

Figure 5.12: [The same as Figure 5.11 but for 12 redshift bins]. The signal-to-noise (eq. (5.64)) computed using simulations, in comparison to the RMS estimate (solid curves), in 12 redshift bins (see figure 5.4). The points represent the standard deviation of the 100 realizations, and the error bars denote the standard error of the standard deviation. These simulations have a resolution of $\lambda_{\text{min}} \sim 484$ Mpc and utilized the default summation bounds ($\ell_{\text{min}} = 3000,\ \ell_{\text{max}} = \infty$) for the signal and noise calculations.

$N_{\rm bin}$	$L_{\rm max}$	$\sum_{\rm bin}\sum_{L}^{L_{\rm max}}(2L+1)$
6	6, 32, 55, 78, 100, 123	36086
12	4, 16, 28, 39, 51, 62, 73, 85, 96, 107, 118, 131	74946
24	3, 8, 14, 20, 25, 31, 37, 42, 48, 54, 59, 65, 71, 76, 82, 88, 93, 99, 105, 110, 116, 121, 130, 132	150853

Table 5.1: Number of modes

modes.

For the kSZ effect considered here, the sum over $2L+1$ is performed in each bin, therefore one might naively guess that it is possible to get $N_{\rm bin}$ times more modes than the primary CMB. However, a more careful estimate needs to be done because the value of $L_{\rm max}$ varies in each bin depending on the size of the signal. Consider the signal-to-noise, described by the ratio of (5.66) and (5.63), for the scales considered here (see eq. (5.65) and use $k_{\rm max} = 278\ H_0$). For $N_{\rm bin}$ bins, in each bin, the value of $L_{\rm max}$ up to which modes can be measured is found by ensuring that

1. For $L < L_{\rm max}$, the signal-to-noise is bigger than 1.
2. For $L < L_{\rm max}$, the signal is dominated by modes larger than 100 Mpc. More precisely, $b_{L0}^{k_{\rm max}} > 0.95 b_{L0}^{\infty}$ for $k_{\rm max} = 278\ H_0$.[3]

Table 5.1 shows the values of $L_{\rm max}$ computed in each bin for the 6, 12 and 24 bin configurations. In every case, it was a failure of criterion (2.) that determined $L_{\rm max}$, as the signal-to-noise is always much bigger than 1 for this range in L. Notice that by doubling the bin size, we approximately double the number of modes. This allows us to extrapolate our results from the three bin configurations.

Figure 5.13 shows how the number of modes increases with $N_{\rm bin}$ based on our estimates using 6, 12 and 24 bins. Extrapolating the data points to higher values of $N_{\rm bin}$ shows that at least 30 bins are needed to match the number of modes in the primary CMB. The same increasing trend should continue until $N_{\rm bin} \sim 50$ (producing 309656 modes), at which point we estimate that the signal-to-noise will drop below 1 in the high redshift bins, thus failing to satisfy criterion (1.), and causing a less rapid increase in the number of modes as a function of $N_{\rm bin}$. Therefore, this method can theoretically achieve more modes than the primary CMB. However, note that within the observable Universe, on scales $\lambda > 100$ Mpc, there are approximately $(\chi_{\rm dec} k_{\rm max}/(2\pi))^3 \sim 2.7 \times 10^6$ total modes. Thus, while remaining competitive

[3]The 95% threshold is arbitrary.

with the primary CMB, the proposed method still falls about an order of magnitude short of the *total* number of possible modes. In addition, because they provide at least partially independent constraints, combining the information from the primary CMB and the information from kSZ tomography can in principle constrain a larger number of modes than either individually.

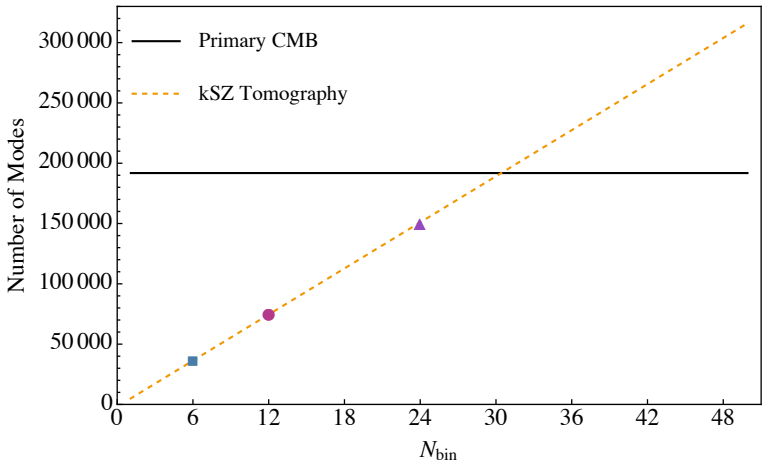

Figure 5.13: The bullet points show the number of modes that can possibly be measured using $N_{\rm bin} = 6$, 12, and 24. Extrapolating these data points (dashed line) indicates that we need at least 30 bins to match the number of modes in the primary CMB at $\ell \leq 437$ (solid line). These calculations were performed with the default summation bounds ($\ell_{\rm min} = 3000,\ \ell_{\rm max} = \infty$).

5.8 Detectability

How close can we get to the cosmic variance limited result with the next generation of CMB experiments and galaxy surveys? Although a complete treatment is beyond the scope of this work, we can give a rough estimate here. The two parameters affecting detectability in the analysis above are the filtering scale $\ell_{\rm min}$ and resolution $\ell_{\rm max}$. The filtering scale is a parameter which can be optimized in any hypothetical analysis. By varying $\ell_{\rm max}$, we can define a rough target for the instrumental noise, resolution, and foreground residuals necessary for a detection to be made.

Previously, we had made the fiducial choice $\ell_{\rm min} = 3000$, which is roughly the angular scale on which the kSZ power surpasses that in the primary CMB. Here, we will explore the filtering scales $\ell_{\rm min} = 2,\ 1000,\ 3000$. In addition, we chose $\ell_{\rm max} \to \infty$, corresponding to the cosmic variance limit. Here, we consider a low-resolution scenario with $\ell_{\rm max} = 3000$ and high-resolution scenario with $\ell_{\rm max} = 5000$.

Starting with the high resolution scenario, in figure 5.14, we show the signal and noise for six redshift bins, choosing $\ell_{\rm min} = 2,\ 1000,\ 3000$ and $\ell_{\rm max} = 5000$. The choice $\ell_{\rm min} = 2$, where

the noise includes the primary CMB, is clearly not optimal as the signal to noise is at most close to one. Raising the filtering scale to $\ell_{\rm min} = 1000$, the signal drops slightly (an effect that is more pronounced at low redshift), but the noise drops by an order of magnitude, raising the signal to noise accordingly. Raising the filtering scale further to $\ell_{\rm min} = 3000$ again further increases the signal to noise. Clearly, removing as much of the primary CMB as possible through filtering is the choice that will optimize signal to noise. Comparing the dotted curves ($\ell_{\rm min} = 3000$) with figure 5.9, reducing $\ell_{\rm max} = \infty$ to $\ell_{\rm max} = 5000$ degrades the signal to noise by about a factor of $10 - 100$, with the degredation more pronounced in the large-redshift bins. Nevertheless, the signal to noise is still $S/N \sim 10 - 100$ over a significant range in L.

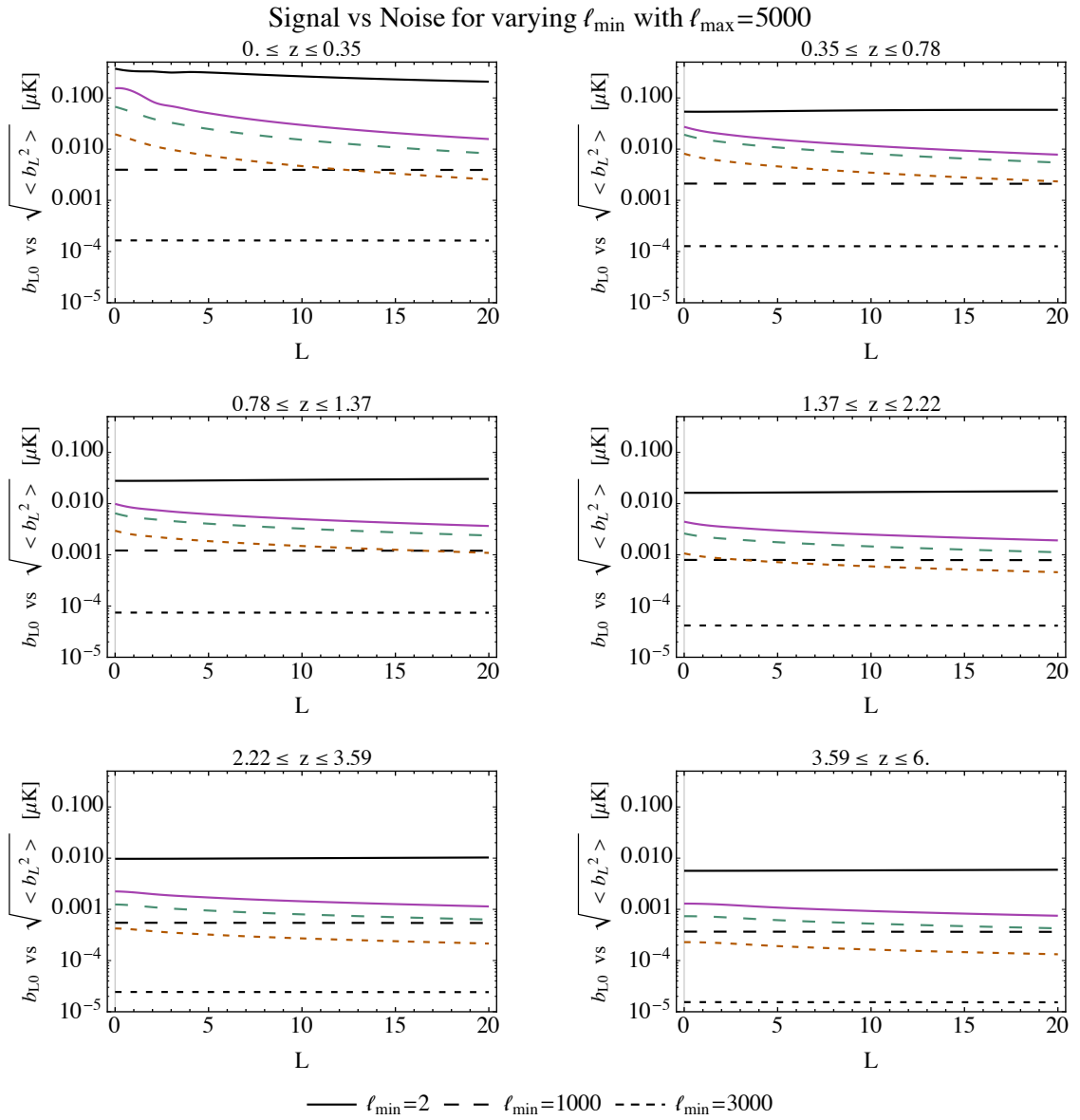

Figure 5.14: The signal (5.66) (coloured lines) and noise (5.63) (black lines) are shown in six redshift bins, computed with $k_{\rm max} = \infty$ and $\ell_{\rm max} = 5000$. We consider different filter scales: $\ell_{\rm min} = 2$ (solid), $\ell_{\rm min} = 1000$ (long dash), $\ell_{\rm min} = 3000$ (short dash).

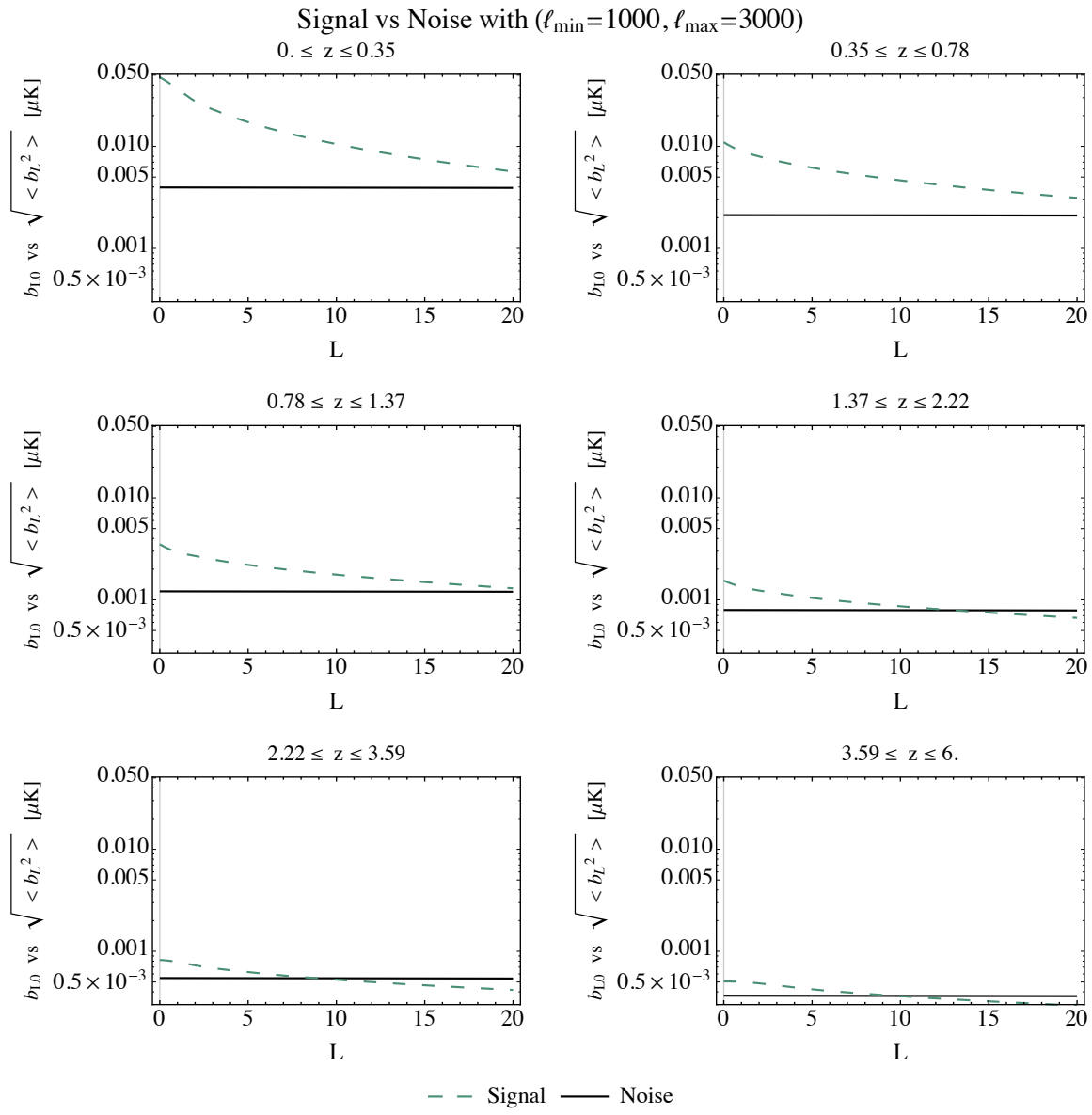

Figure 5.15: The signal (5.66) (dashed) and noise (5.63) (solid) are shown in six redshift bins, computed with $k_{\max} = \infty$, $\ell_{\min} = 1000$, and $\ell_{\max} = 3000$.

Moving on to the low resolution scenario, in figure 5.15 we show the signal and noise for six redshift bins, choosing $\ell_{\rm min} = 1000$ and $\ell_{\rm max} = 3000$. Here, it can be seen that the signal-to-noise is significant only for the low redshift bins, and reaches at most $S/N \sim 10$.

With these results, we can map the low and high resolution scenarios to a rough set of experimental requirements. In particular, we consider CMB instrumental noise, foreground residuals, finite resolution, and galaxy shot noise. The treatment of a number of important systematics is beyond our scope. We assume that the instrumental noise contribution to the CMB temperature is gaussian and uniform on the sky,[3] gaussian beams, and that foreground residuals in the cleaned data product used for the analysis can be modelled as a uniform gaussian random field. Under these assumptions, the measured CMB temperature power spectrum is modelled as:

$$C_\ell^{TT} = \left(C_\ell^{TT,\rm kSZ} + C_\ell^{TT,\rm p} + N_\ell^{\rm CMB} + F_\ell^{\rm CMB}\right) \exp\left[\frac{\ell(\ell+1)\theta^2_{\rm FWHM}}{8\ln 2}\right], \tag{5.67}$$

where $\theta_{\rm FWHM}$ is the full-width at half-maximum of the Gaussian beam in radians. The noise and foreground contributions are

$$N_\ell^{\rm CMB} = (\sigma_N\ \theta_{\rm FWHM})^2, \tag{5.68}$$

where σ_N is the noise per variance in each beam-sized patch, and

$$F_\ell^{\rm CMB} = (\sigma_F\ \theta_{\rm FWHM})^2, \tag{5.69}$$

where σ_F is the variance in foreground residuals in each beam-sized patch.

Assuming that galaxy number density is an unbiased tracer of free electrons,[4] and assuming that redshift bins are far larger than redshift errors, the measurement of the density angular power spectrum is limited by shot noise, yielding:

$$C_{\ell'}^{\delta\delta}(\bar\chi_e) = C_\ell^{\rm gg}(\bar\chi_e) + N_\ell^{\rm gg}(\bar\chi_e), \tag{5.70}$$

with

$$N_\ell^{\rm gg}(\bar\chi_e) = \frac{1}{N_g(\bar\chi_e)}, \tag{5.71}$$

where N_g is the number of galaxies per square radian in a redshift bin centred on $\bar\chi_e$.

We collect the experimental requirements in table 5.2 for the case of 6 redshift bins. We set the resolution by requiring that the effect of the beam is not dominant, yielding $\theta_{\rm FWHM} = (8\ln 2/\ell_{\rm max}(\ell_{\rm max}+1))^{1/2}$. The requirements on the CMB instrumental noise and foregrounds are then set by the condition that one measures the CMB power spectrum at

[3] This is clearly incorrect, and spatially varying noise will be an important systematic to assess in the future.

[4] We reserve a more complete treatment for future work.

$\ell_{\rm max}$	$\theta_{\rm FWHM}$ (arcmin)	$\sigma_N,\ \sigma_F$ (μK $\theta^{-1}_{\rm FWHM}$)	$N_g(\bar{\chi}_e)$ (arcmin^{-2})
3000	2.7	3.7	14, 26, 82, 255, 697, 1661
5000	1.6	3.0	45, 62, 155, 500, 1700, 5000

Table 5.2: The experimental characteristics required to access the low-resolution ($\ell_{\rm max} = 3000$) and high resolution ($\ell_{\rm max} = 5000$) scenarios.

a signal to noise of one at $\ell_{\rm max}$, e.g. $N^{\rm CMB}_{\ell_{\rm max}}, F^{\rm CMB}_{\ell_{\rm max}} = C^{TT,{\rm kSZ}}_{\ell_{\rm max}} + C^{TT,{\rm p}}_{\ell_{\rm max}}$. Finally, we solve for $N_g(\bar{\chi}_e)$ from the condition that one measures the galaxy-galaxy power spectrum with signal to noise of one, e.g. $C^{\rm gg}_{\ell_{\rm max}}(\bar{\chi}_e) = N^{\rm gg}_{\ell_{\rm max}}(\bar{\chi}_e)$.

Our choices for resolution correspond roughly to the range of resolutions considered for Stage 4 CMB experiments [246] of $\theta_{\rm FWHM} \simeq 1 - 3$ arcmin. The requirement on instrumental noise for the two resolution scenarios we consider falls within the target of $\sigma_N \theta_{\rm FWHM} \sim 1\mu$K arcmin^{-1} [246]. It is expected that foreground residuals will not dominate the signal until $\ell \gtrsim 3000$ [246], making it unlikely that the high resolution scenario could be achieved without better frequency coverage, as could be attained from space. The forecasted galaxy number density is $N_g \sim 30$ for Euclid [247], and about $N_g \sim 130$ for LSST [233]. This will be sufficient to recover information from the first 2-3 redshift bins (which covers most of the reach of such surveys) in either resolution scenario; a more optimal binning strategy can be constructed for a specific survey. It is unlikely that the higher redshift bins could be accessed with a galaxy survey. For comparison, the galaxy density in the Hubble Deep Field [248] is $N_g \sim 500$, and $N_g \sim 1700$ in the Hubble Ultra Deep Field [249] (which is close to the cosmic variance limit). Since these surveys are fairly complete to high redshift, it is likely that bins 5 and 6 in the high resolution case require more galaxies than there are in the observable Universe. It may be possible to access the high resolution example with new techniques such as intensity mapping.

In conclusion, it seems that a first detection could in principle be made with the next generation of CMB experiments and galaxy surveys. Based on the rough analysis above, future progress can be made with increases in resolution of CMB experiments, better foreground subtraction using multifrequency information, and new techniques for measuring the angular matter spectrum to high resolution, such as high resolution intensity mapping. A more detailed forecast will be presented in future work.

5.9 Discussion and Conclusions

In this chapter, we have assessed the ability of kSZ tomography to yield information about the long-wavelength Universe in the cosmic variance limit. The signal of interest is a power asymmetry in the direct cross correlation of the kSZ contribution to the CMB temperature and the electron density binned at various redshifts. We quantify this signal in terms of power multipoles, and compared it with the amplitude of "accidental" power asymmetry due to the statistically homogeneous component of the kSZ contribution to the CMB temperature (the primary source of cosmic variance in this context). The results are promising in this highly optimistic scenario, yielding a signal to noise greater than unity over a large range of power multiples in a large number of redshift bins. A first forecast indicates that next-generation CMB experiments and galaxy surveys should be able to make a detection at low redshift and large angular scale.

Although we have established that there is in principle a signal to detect, there is a significant amount of work that must be done to assess what can be done in practice. First, a more detailed forecast should be done to determine what is required to reach the necessary threshold in sensitivity, accuracy, and resolution. Next, there are important systematic errors and other potential sources of power asymmetry that should be investigated including, but not limited to: relativistic aberration of the CMB [250] and large scale structure [251, 252], clustering [253], electron bias in the nonlinear regime (e.g. [254]), non-Gaussian aspects of CMB lensing [255], redshift space distortions [256], asymmetric scan strategies (e.g. [257]) and incomplete sky coverage, incomplete LSS surveys, asymmetric beams and point spread functions, and more realistic window functions. Furthermore, the signal may be boosted in the presence of some types of primordial non-Gaussianity [258]. We leave a more careful investigation of these and other important aspects to future work.

If it is indeed possible to approach the cosmic variance limited scenario we have described in this chapter, what would we stand to learn? Because kSZ tomography is probing different portions of the surface of last scattering than the primary CMB, the constraining power for various early Universe scenarios involving extra sources of inhomogeneity can be extraordinary. For example, in Ref. [113] it was shown that the constraints on parameters in a theory that predicts cosmic bubble collisions could improve by several orders of magnitude comparing kSZ tomography to existing and forecasted constraints from the primary CMB. One might expect constraints on theories of the large-scale CMB anomalies to be similarly impressive. Given a high enough fidelity measurement, it should also be possible to reconstruct the 3D large-scale gravitational potential throughout much of the observable Universe (in analogy with

Ref. [259]). Such a reconstruction could be an important tool, for example in future studies of primordial non-Gaussianity. Performing this exercise would also significantly clarify precisely what new information, beyond that encoded in the primary CMB and CMB polarization, there is to gain from kSZ tomography. In any case, we stand to learn a great deal about cosmology in the coming era of precision measurements of CMB secondaries, with the kSZ effect playing a leading role. In this new era, it is important to understand the nuances of the kSZ effect, and to target new observables.

Chapter 6

Conclusions

It may seem unfathomable to attempt constructing a physical model of the entire large scale universe. Remarkably, decades of theoretical and observational advances have converged on a simple standard cosmological model, ΛCDM. The current state of the model is outlined in chapter 1. Despite its many successes, there are several aspects of the theory that we fail to understand. The focus for this dissertation is infrared cosmology, the regime in which the mystery of dark energy is at the top of list.

In attempting to understand dark energy and accelerated expansion, we are led to question the regime of validity of GR. The successes of general relativity on solar system scales do not necessarily imply that the theory should be valid on the enormous scales we consider in cosmology. All evidence for dark energy may instead be evidence that GR is breaking down on cosmological scales. Infrared modifications to GR can potentially offer a better explanation for cosmic expansion. This dissertation has focused on *massive gravity*, a subset of modified gravity theories in which the graviton has a tiny non-zero mass. Important theoretical aspects of the theory are described in chapter 2.

Massive gravity is referred to as an infrared modification to GR, so it's important that it is exactly that: there should only be modification on large scales. On smaller scales, we are very happy with the predictions of GR, so we don't want the model's new features to spoil that. In massive gravity, the mechanism that restores predictions of GR on small scales is the Vainshtein screening mechanism. The contributions presented in chapter 3 analyzed the fully non-linear dynamical behaviour of the screening mechanism in two related infrared modifications: the DGP model and dRGT massive gravity. Our work is within the decoupling limit of these models, in which a scalar galileon field contains all relevant information about the modification and screening mechanism. By numerically solving the equation of motion for this

scalar field we showed that for a wide range of generic initial conditions, the screening solutions around a static source are accessed as the end state of the evolution. We also incorporated toy models for collapsing and exploding sources, showing that the screening mechanism adapts as expected throughout the evolution.

However, the evolution is only well posed for sufficiently low density sources, and sufficiently long collapse/explosion time scales. We discovered that for real astrophysical parameters, the evolution would enter the strong coupling regime of the theory. At this point, the classical effective field theory can not be trusted because perturbativity breaks down, meaning that quantum corrections are required to determine the future evolution. Here, we see that the lack of a UV completion hinders our ability to make useful predictions within this model. Whether or not there exists a standard (or even non-standard) UV completion for a massive graviton is unknown. Since the strong coupling scale Λ_3 is still rather low, one might just hope for a partial UV completion that could raise the cutoff to the Planck scale.

Extending massive gravity to bigravity leads to a theory with two dynamical spin-2 fields that interact through the dRGT mass term. In this model, two viable classes of FRW background solutions exist, allowing us to compare bigravity's predictions for cosmology to the standard ΛCDM model. To further test these candidate background solutions, we must fully analyze the perturbations around them. In chapter 4, the focus was on the tensor sector of perturbations. Gravitational waves are quite interesting in bigravity since there are physical and dark perturbations that interact in a non-trivial manner.

In one class of background solutions, the "expanding branch", there was no significant deviation in the behaviour of physical gravitational waves compared to standard ΛCDM. In the "bouncing branch" background, we found that gravitational waves exhibit power-law growth with time. Since generically, amplification of perturbations could be enough to rule out this branch of the model, understanding this growth was our goal. We carefully analyzed how the growth of gravitational waves was sensitive to parameters of the theory, specifically, the initial conditions of the physical and dark gravitational waves.

It was therefore necessary to address the issue of initial conditions by computing the inflationary power spectra for both sets of tensor perturbations. We found that inflation sets an enormous suppression of the dark gravitational waves, resulting in physical gravitational waves that are not observably distinct from that of GR. Although we did not find evidence for any compelling deviation from standard cosmology in this analysis, fully understanding the predictions of these models is imperative, as small but observable differences can hold the key to testing GR on cosmological scales.

Together, chapters 3 and 4 considered two central aspects of modified gravity theories: agreement with GR on small scales and deviation from GR on cosmological scales. There seems to be a general trend that if a model does not posses a screening mechanism to match GR on small scales, it is ruled out; if the model does have a limit that looks like GR, it gets pushed into this limit, becoming observationally identical to GR. Tracking down potential ways to test these models is challenging, and cosmologists are actively pursuing this challenge [232, 31, 30, 260, 261, 262, 263, 233].

Even if they lack a testable distinction with present technology, these models offer an explanation for cosmic acceleration without a cosmological constant, providing a different interpretation of dark energy and the cosmological constant problem. With that being said, many would say that massive gravity and its extensions still fail to solve the "old" cosmological constant problem of why the cosmological constant from the vacuum does not gravitate. These models have to assume that some unknown mechanism exists to set the expected large cosmological constant to zero, but perhaps this is still an improvement over setting it to a specific unnatural finely tuned value. Furthermore, massive gravity offers a technically natural explanation for the small parameter of the theory that drives accelerated expansion.

Efforts in constructing viable models and understanding their predictions should be matched by efforts in examining observational probes that will allow us to constrain, validate, or rule out these models. Searching for deviations of GR and ΛCDM on very large scales leads us to look out to the edge of our observable horizon captured in the CMB. The CMB's ability to constrain cosmology cannot be overstated, and yet, there is an inherent limit on the precision with which we can probe scales on the order of our horizon: *cosmic variance*. In fact, we cannot learn anything further about these large scales from the primary CMB alone. How can we do better?

With so many experiments being proposed, it is necessary to compare the constraining power of different probes of large scales. The ultimate goal is to unlock the most information possible. Chapter 5 makes contributions on this front by considering a novel probe of large scales, the kSZ effect, which arises from Compton scattering of CMB photons with moving free electrons. The intriguing aspect of this effect lies in its ability to mitigate cosmic variance: the signal is proportional to the observed CMB dipole at the location of each electron. Since each electron in the universe probes a different portion of the surface of last scattering, detection of the kSZ effect can yield more information than the cosmic variance limited measurements we can make from one position here on Earth.

This work lays the theoretical foundation for the large-scale kSZ effect, computing the signal-to-noise in the cosmic variance limited (noise-free) scenario. In this limit, we found that

the signal-to-noise is huge, establishing that there is indeed a real effect that can in principle be detected. We also presented a forecast to indicate that putting together next-generation CMB experiments and galaxy surveys we may be able to detect the large-scale kSZ signal at low redshift. These first steps of analysis pave the way for future work in more thoroughly assessing the realistic detectability of the effect, including systematic errors. Also important is understanding the usefulness of a detection in terms of constraining power for various models with large-scale inhomogeneities, in comparison to the information already provided in the CMB.

Solidifying our knowledge of theoretical aspects and predictions of modified gravity theories is essential to illuminate the best basis for models of our universe. Even just within the class of massive gravity, we have learned so much about what is required for an acceptable viable theory of gravity. Beyond theoretical consistency, any model must be consistent with the plethora of current cosmological and astrophysical data. This dataset will certainly grow considering the numerous exciting avenues of future observational endeavours. It is crucial to understand what novel observational probes, like the kSZ effect, can tell us about the remaining mysteries of the universe.

References

[1] R. Brito, A. Terrana, M. C. Johnson and V. Cardoso, *Nonlinear dynamical stability of infrared modifications of gravity*, *Phys. Rev. D* **90** (Dec., 2014) 124035, [1409.0886].

[2] M. C. Johnson and A. Terrana, *Tensor modes in bigravity: Primordial to present*, *Phys. Rev. D* **92** (Aug., 2015) 044001, [1503.05560].

[3] A. Terrana, M.-J. Harris and M. C. Johnson, *Analyzing the cosmic variance limit of remote dipole measurements of the cosmic microwave background using the large-scale kinetic sunyaev zel'dovich effect*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) 040.

[4] Planck Collaboration, R. Adam, P. A. R. Ade, N. Aghanim, Y. Akrami, M. I. R. Alves et al., *Planck 2015 results. I. Overview of products and scientific results*, *Astronomy and Astrophysics* **594** (Sept., 2016) A1, [1502.01582].

[5] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson et al., *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems*, *Astronomical Journal* **142** (Sept., 2011) 72, [1101.1529].

[6] S. Carroll, *Spacetime and Geometry: An Introduction to General Relativity*. Benjamin Cummings, 2003.

[7] S. Dodelson, *Modern cosmology*. Academic Press, San Diego, CA, 2003.

[8] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont et al., *Planck 2015 results. XIII. Cosmological parameters*, *ArXiv e-prints* (Feb., 2015) , [1502.01589].

[9] A. G. Riess, L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, A. V. Filippenko et al., *A 2.4% Determination of the Local Value of the Hubble Constant*, *Astrophysical Journal* **826** (July, 2016) 56, [1604.01424].

[10] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich et al., *Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant*, *Astronomical Journal* **116** (Sept., 1998) 1009–1038, [astro-ph/9805201].

[11] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro et al., *Measurements of* Ω *and* Λ *from 42 High-Redshift Supernovae*, *Astrophysical Journal* **517** (June, 1999) 565–586, [astro-ph/9812133].

[12] A. Rest, D. Scolnic, R. J. Foley, M. E. Huber, R. Chornock, G. Narayan et al., *Cosmological Constraints from Measurements of Type Ia Supernovae Discovered during the First 1.5 yr of the Pan-STARRS1 Survey*, *Astrophysical Journal* **795** (Nov., 2014) 44, [1310.3828].

[13] A. D. Linde, *A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems*, *Phys. Lett.* **B108** (1982) 389–393.

[14] A. Albrecht and P. J. Steinhardt, *Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking*, *Phys. Rev. Lett.* **48** (1982) 1220–1223.

[15] A. Linde, *Inflationary Cosmology after Planck 2013*, in *Proceedings, 100th Les Houches Summer School: Post-Planck Cosmology: Les Houches, France, July 8 - August 2, 2013*, pp. 231–316, 2015. 1402.0526. DOI.

[16] C. P. Burgess, M. Cicoli and F. Quevedo, *String Inflation After Planck 2013*, *JCAP* **1311** (2013) 003, [1306.3512].

[17] R. Penrose, *Difficulties with inflationary cosmology*, *Annals N. Y. Acad. Sci.* **571** (1989) 249–264.

[18] J. Khoury, B. A. Ovrut, P. J. Steinhardt and N. Turok, *The Ekpyrotic universe: Colliding branes and the origin of the hot big bang*, *Phys. Rev.* **D64** (2001) 123522, [hep-th/0103239].

[19] A. Borde, A. H. Guth and A. Vilenkin, *Inflationary space-times are incompletein past directions*, *Phys. Rev. Lett.* **90** (2003) 151301, [gr-qc/0110012].

[20] G. W. Gibbons and N. Turok, *The Measure Problem in Cosmology*, *Phys. Rev.* **D77** (2008) 063516, [hep-th/0609095].

[21] A. Ijjas, P. J. Steinhardt and A. Loeb, *Inflationary paradigm in trouble after Planck2013*, *Phys. Lett.* **B723** (2013) 261–266, [1304.2785].

[22] A. Fertig, *Alternatives to Inflation non-minimal ekpyrosis and conflation*. PhD thesis, Humboldt U., Berlin, 2016.

[23] V. Mukhanov, *Physical Foundations of Cosmology*. Cambridge Univ. Press, Cambridge, 2005.

[24] M. Roos, *Dark Matter: The evidence from astronomy, astrophysics and cosmology*, *ArXiv e-prints* (Jan., 2010) , [1001.0316].

[25] M. S. Turner, *Windows on the Axion*, *Phys. Rept.* **197** (1990) 67–97.

[26] L. D. Duffy and K. van Bibber, *Axions as Dark Matter Particles*, *New J. Phys.* **11** (2009) 105008, [0904.3346].

[27] A. H. G. Peter, *Dark Matter: A Brief Review*, *ArXiv e-prints* (Jan., 2012) , [1201.3942].

[28] L. Bergstrom, *Dark Matter Evidence, Particle Physics Candidates and Detection Methods*, *Annalen Phys.* **524** (2012) 479–496, [1205.4882].

[29] M. Klasen, M. Pohl and G. Sigl, *Indirect and direct search for dark matter*, *Prog. Part. Nucl. Phys.* **85** (2015) 1–32, [1507.03800].

[30] Planck collaboration, P. A. R. Ade et al., *Planck 2015 results. XIV. Dark energy and modified gravity*, *Astron. Astrophys.* **594** (2016) A14, [1502.01590].

[31] DES collaboration, T. Abbott et al., *The Dark Energy Survey: more than dark energy - an overview*, *Mon. Not. Roy. Astron. Soc.* (2016) , [1601.00329].

[32] R. Bousso, *The Cosmological Constant Problem, Dark Energy, and the Landscape of String Theory*, *ArXiv e-prints* (Mar., 2012) , [1203.0307].

[33] S. M. Carroll, *The Cosmological Constant*, *Living Reviews in Relativity* **4** (Dec., 2001) 1, [astro-ph/0004075].

[34] N. Seiberg, *Naturalness versus supersymmetric nonrenormalization theorems*, *Phys. Lett.* **B318** (1993) 469–475, [hep-ph/9309335].

[35] S. Weinberg, *Anthropic bound on the cosmological constant*, *Phys. Rev. Lett.* **59** (Nov, 1987) 2607–2610.

[36] M. Doran and C. Wetterich, *Quintessence and the cosmological constant*, *Nuclear Physics B Proceedings Supplements* **124** (July, 2003) 57–62, [astro-ph/0205267].

[37] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley et al., *Observation of Gravitational Waves from a Binary Black Hole Merger*, *Physical Review Letters* **116** (Feb., 2016) 061102, [1602.03837].

[38] T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, *Modified Gravity and Cosmology*, *Phys. Rept.* **513** (2012) 1–189, [1106.2476].

[39] A. Joyce, B. Jain, J. Khoury and M. Trodden, *Beyond the Cosmological Standard Model*, 1407.0059.

[40] K. Hinterbichler, *Theoretical Aspects of Massive Gravity*, *Rev.Mod.Phys.* **84** (2012) 671–710, [1105.3735].

[41] C. de Rham, *Massive Gravity*, 1401.4173.

[42] M. Fierz and W. Pauli, *On relativistic wave equations for particles of arbitrary spin in an electromagnetic field*, *Proc.Roy.Soc.Lond.* **A173** (1939) 211–232.

[43] C. de Rham, G. Gabadadze and A. J. Tolley, *Resummation of Massive Gravity*, *Phys.Rev.Lett.* **106** (2011) 231101, [1011.1232].

[44] C. de Rham and G. Gabadadze, *Generalization of the Fierz-Pauli Action*, *Phys.Rev.* **D82** (2010) 044020, [1007.0443].

[45] C. de Rham, J. T. Deskins, A. J. Tolley and S.-Y. Zhou, *Graviton Mass Bounds*, *Rev. Mod. Phys.* **89** (2017) 025004, [1606.08462].

[46] C. Talmadge, J. P. Berthias, R. W. Hellings and E. M. Standish, *Model-independent constraints on possible modifications of newtonian gravity*, *Phys. Rev. Lett.* **61** (Sep, 1988) 1159–1162.

[47] C. M. Will, *Bounding the mass of the graviton using gravitational wave observations of inspiralling compact binaries*, *Phys. Rev.* **D57** (1998) 2061–2068, [gr-qc/9709011].

[48] Virgo, LIGO Scientific collaboration, B. P. Abbott et al., *Tests of general relativity with GW150914*, *Phys. Rev. Lett.* **116** (2016) 221101, [1602.03841].

[49] G. Dvali, A. Gruzinov and M. Zaldarriaga, *The Accelerated universe and the moon*, *Phys. Rev.* **D68** (2003) 024012, [hep-ph/0212069].

[50] A. Lue and G. Starkman, *Gravitational leakage into extra dimensions: Probing dark energy using local gravity*, *Phys. Rev.* **D67** (2003) 064002, [astro-ph/0212083].

[51] A. Gruzinov, *On the graviton mass*, *New Astron.* **10** (2005) 311–314, [astro-ph/0112246].

[52] D. J. Fixsen, *The Temperature of the Cosmic Microwave Background*, *Astrophysical Journal* **707** (Dec., 2009) 916–920, [0911.1955].

[53] A. A. Penzias and R. W. Wilson, *A Measurement of Excess Antenna Temperature at 4080 Mc/s.*, *Astrophysical Journal* **142** (July, 1965) 419–421.

[54] G. F. Smoot, L. Tenorio, A. J. Banday, A. Kogut, E. L. Wright, G. Hinshaw et al., *Statistics and Topology of the COBE DMR First Year Maps*, *Astrophys. J.* **437** (1994) 1–11, [astro-ph/9312031].

[55] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley et al., *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, *The Astrophysical Journal Supplement* **208** (Oct., 2013) 19, [1212.5226].

[56] Planck collaboration, P. A. R. Ade et al., *Planck 2013 results. I. Overview of products and scientific results*, *Astron. Astrophys.* **571** (2014) A1, [1303.5062].

[57] Planck collaboration, P. A. R. Ade et al., *Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **571** (2014) A24, [1303.5084].

[58] W. Hu and N. Sugiyama, *Toward understanding CMB anisotropies and their implications*, *Phys. Rev.* **D51** (1995) 2599–2630, [astro-ph/9411008].

[59] W. Hu, N. Sugiyama and J. Silk, *The Physics of microwave background anisotropies*, *Nature* **386** (1997) 37–43, [astro-ph/9604166].

[60] A. Challinor and H. Peiris, *Lecture notes on the physics of cosmic microwave background anisotropies*, in *American Institute of Physics Conference Series* (M. Novello and S. Perez, eds.), vol. 1132 of *American Institute of Physics Conference Series*, pp. 86–140, May, 2009. 0903.5158. DOI.

[61] W. Hu and M. J. White, *A CMB polarization primer*, *New Astron.* **2** (1997) 323, [astro-ph/9706147].

[62] R. K. Sachs and A. M. Wolfe, *Perturbations of a Cosmological Model and Angular Variations of the Microwave Background*, *Astrophysical Journal* **147** (Jan., 1967) 73.

[63] C. Carbone, M. Baldi, V. Pettorino and C. Baccigalupi, *Maps of CMB lensing deflection from N-body simulations in Coupled Dark Energy Cosmologies*, *JCAP* **1309** (2013) 004, [1305.0829].

[64] L. Amendola, V. Pettorino, C. Quercellini and A. Vollmer, *Testing coupled dark energy with next-generation large-scale observations*, *Phys. Rev. D* **85** (May, 2012) 103008, [1111.1404].

[65] M. Baldi and V. Pettorino, *High-z massive clusters as a test for dynamical coupled dark energy*, *Monthly Notices of the RAS* **412** (Mar., 2011) L1–L5, [1006.3761].

[66] L. Amendola, G. Ballesteros and V. Pettorino, *Effects of modified gravity on B-mode polarization*, *Phys.Rev.* **D90** (2014) 043009, [1405.7004].

[67] N. Aghanim, S. Majumdar and J. Silk, *Secondary anisotropies of the CMB*, *Rept. Prog. Phys.* **71** (2008) 066902, [0711.0518].

[68] D. J. Schwarz, C. J. Copi, D. Huterer and G. D. Starkman, *CMB Anomalies after Planck*, *Class. Quant. Grav.* **33** (2016) 184001, [1510.07929].

[69] E. M. George et al., *A measurement of secondary cosmic microwave background anisotropies from the 2500-square-degree SPT-SZ survey*, *Astrophys. J.* **799** (2015) 177, [1408.3161].

[70] N. Sehgal, H. Trac, V. Acquaviva, P. A. R. Ade, P. Aguirre, M. Amiri et al., *The Atacama Cosmology Telescope: Cosmology from Galaxy Clusters Detected via the Sunyaev-Zel'dovich Effect*, *Astrophysical Journal* **732** (May, 2011) 44, [1010.1025].

[71] C. Heymans, L. Van Waerbeke, L. Miller, T. Erben, H. Hildebrandt, H. Hoekstra et al., *CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey*, *Monthly Notices of the RAS* **427** (Nov., 2012) 146–166, [1210.0032].

[72] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont et al., *Planck early results. VIII. The all-sky early Sunyaev-Zeldovich cluster sample*, *Astronomy and Astrophysics* **536** (Dec., 2011) A8, [1101.2024].

[73] Astrophysics Group, Cavendish Laboratory, Cambridge, UK collaboration, A. C. J. T. L. Zwart et al., *The Arcminute Microkelvin Imager*, *Mon. Not. Roy. Astron. Soc.* **391** (2008) 1545, [0807.2469].

[74] K. S. Dawson, W. L. Holzapfel, J. E. Carlstrom, M. Joy and S. J. LaRoque, *Final results from the bima cmb anistropy survey and search for signature of the sz effect*, *Astrophys. J.* **647** (2006) 13–24, [astro-ph/0602413].

[75] J. Bock et al., *The cosmic infrared background experiment*, *New Astron. Rev.* **50** (2006) 215–220, [astro-ph/0510587].

[76] W. Hu and S. Dodelson, *Cosmic microwave background anisotropies*, *Ann. Rev. Astron. Astrophys.* **40** (2002) 171–216, [astro-ph/0110414].

[77] M. J. Rees and D. W. Sciama, *Large scale Density Inhomogeneities in the Universe*, *Nature* **217** (1968) 511–516.

[78] A. Lewis and A. Challinor, *Weak gravitational lensing of the cmb*, *Phys. Rept.* **429** (2006) 1–65, [astro-ph/0601594].

[79] Y. B. Zeldovich and R. A. Sunyaev, *The Interaction of Matter and Radiation in a Hot-Model Universe*, *Astrophysics and Space Science* **4** (July, 1969) 301–316.

[80] R. A. Sunyaev and I. B. Zeldovich, *The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement*, *MNRAS* **190** (Feb., 1980) 413–420.

[81] N. Hand, G. E. Addison, E. Aubourg, N. Battaglia, E. S. Battistelli, D. Bizyaev et al., *Evidence of Galaxy Cluster Motions with the Kinematic Sunyaev-Zel'dovich Effect*, *Physical Review Letters* **109** (July, 2012) 041101, [1203.4219].

[82] F. De Bernardis et al., *Detection of the pairwise kinematic Sunyaev-Zel'dovich effect with BOSS DR11 and the Atacama Cosmology Telescope*, 1607.02139.

[83] DES, SPT collaboration, B. Soergel et al., *Detection of the kinematic Sunyaev-Zel'dovich effect with DES Year 1 and SPT*, *Mon. Not. Roy. Astron. Soc.* (2016) , [1603.03904].

[84] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, E. Aubourg et al., *Planck intermediate results. XXXVII. Evidence of unbound gas from the kinetic Sunyaev-Zeldovich effect*, *Astronomy & Astrophysics* **586** (Feb., 2016) A140, [1504.03339].

[85] E. Calabrese et al., *Precision Epoch of Reionization studies with next-generation CMB experiments*, *JCAP* **1408** (2014) 010, [1406.4794].

[86] W. L. K. Wu, J. Errard, C. Dvorkin, C. L. Kuo, A. T. Lee, P. McDonald et al., *A Guide to Designing Future Ground-based Cosmic Microwave Background Experiments*, *Astrophys. J.* **788** (2014) 138, [1402.4108].

[87] S. Ho, S. Dedeo and D. Spergel, *Finding the Missing Baryons Using CMB as a Backlight*, *ArXiv e-prints* (Mar., 2009) , [0903.2845].

[88] J. Shao, P. Zhang, W. Lin, Y. Jing and J. Pan, *Kinetic Sunyaev-Zel'dovich tomography with spectroscopic redshift surveys*, *MNRAS* **413** (May, 2011) 628–642, [1004.1301].

[89] P. Zhang and A. Stebbins, *Confirmation of the Copernican Principle at Gpc Radial Scale and above from the Kinetic Sunyaev-Zel'dovich Effect Power Spectrum*, *Physical Review Letters* **107** (July, 2011) 041301, [1009.3967].

[90] P. Zhang and U.-L. Pen, *Deprojecting Sunyaev-Zeldovich Statistics*, *Astrophysical Journal* **549** (Mar., 2001) 18–27, [astro-ph/0007462].

[91] D. Munshi, I. T. Iliev, K. L. Dixon and P. Coles, *Extracting the late-time kinetic Sunyaev-Zel'dovich effect*, 1511.03449.

[92] E. Schaan, S. Ferraro, M. Vargas-Magaña, K. M. Smith, S. Ho, S. Aiola et al., *Evidence for the kinematic Sunyaev-Zel'dovich effect with the Atacama Cosmology Telescope and velocity reconstruction from the Baryon Oscillation Spectroscopic Survey*, *Phys. Rev. D* **93** (Apr., 2016) 082002.

[93] S. Ferraro, J. C. Hill, N. Battaglia, J. Liu and D. N. Spergel, *The Kinematic Sunyaev-Zel'dovich Effect with Projected Fields II: prospects, challenges, and comparison with simulations*, 1605.02722.

[94] J. C. Hill, S. Ferraro, N. Battaglia, J. Liu and D. N. Spergel, *Kinematic Sunyaev-Zel'dovich Effect with Projected Fields: A Novel Probe of the Baryon Distribution with Planck, WMAP, and WISE Data*, *Phys. Rev. Lett.* **117** (2016) 051301, [1603.01608].

[95] H. Miao, W.-P. Lin and P.-J. Zhang, *Probing the Missing Baryons via kSZ Stacking*, *Research in Astronomy and Astrophysics* **16** (Apr., 2016) 015.

[96] C. Hernández-Monteagudo, Y.-Z. Ma, F. S. Kitaura, W. Wang, R. Génova-Santos, J. Macías-Pérez et al., *Evidence of the Missing Baryons from the Kinematic Sunyaev-Zeldovich Effect in Planck Data*, *Physical Review Letters* **115** (Nov., 2015) 191301, [1504.04011].

[97] X.-d. Xu, B. Wang and P. Zhang, *Testing Tensor-Vector-Scalar Theory with latest cosmological observations*, *ArXiv e-prints* (Dec., 2014) , [1412.4073].

[98] E.-M. Mueller, F. de Bernardis, R. Bean and M. Niemack, *Constraints on gravity and dark energy from the pairwise kinematic Sunyaev-Zeldovich effect*, *ArXiv e-prints* (Aug., 2014) , [1408.6248].

[99] F. Bianchini and A. Silvestri, *Kinetic Sunyaev-Zel'dovich effect in modified gravity*, *Phys. Rev.* **D93** (2016) 064026, [1510.08844].

[100] A. Kashlinsky, F. Atrio-Barandela, D. Kocevski and H. Ebeling, *A Measurement of Large-Scale Peculiar Velocities of Clusters of Galaxies: Results and Cosmological Implications*, *The Astrophysical Journal, Letters* **686** (Oct., 2008) L49–L52, [0809.3734].

[101] P. Zhang, *The dark flow induced small-scale kinetic Sunyaev-Zel'dovich effect*, *MNRAS* **407** (Sept., 2010) L36–L40, [1004.0990].

[102] A. Kashlinsky, F. Atrio-Barandela and H. Ebeling, *Measuring the Dark Flow with Public X-ray Cluster Data*, *Astrophysical Journal* **732** (May, 2011) 1, [1012.3214].

[103] Z. Li, P. Zhang and X. Chen, *Anomalous Anisotropic Cross-correlations between WMAP CMB Maps and SDSS Galaxy Distribution and Implications on the Dark Flow Scenario*, *Astrophysical Journal* **758** (Oct., 2012) 130, [1209.0520].

[104] G. Lavaux, N. Afshordi and M. J. Hudson, *First measurement of the bulk flow of nearby galaxies using the cosmic microwave background*, *MNRAS* **430** (Apr., 2013) 1617–1635, [1207.1721].

[105] F. Atrio-Barandela, A. Kashlinsky, H. Ebeling, D. J. Fixsen and D. Kocevski, *Probing the Dark Flow signal in WMAP 9 yr and PLANCK cosmic microwave background maps*, *ArXiv e-prints* (Nov., 2014) , [1411.4180].

[106] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont et al., *Planck intermediate results. XIII. Constraints on peculiar velocities*, *Astronomy & Astrophysics* **561** (Jan., 2014) A97, [1303.5090].

[107] E.-M. Mueller, F. de Bernardis, R. Bean and M. D. Niemack, *Constraints on Gravity and Dark Energy from the Pairwise Kinematic Sunyaev-Zel'dovich Effect*, *Astrophysical Journal* **808** (July, 2015) 47, [1408.6248].

[108] X.-D. Xu, B. Wang, P. Zhang and F. Atrio-Barandela, *The effect of dark matter and dark energy interactions on the peculiar velocity field and the kinetic Sunyaev-Zel'dovich effect*, *JCAP* **12** (Dec., 2013) 1, [1308.1475].

[109] E.-M. Mueller, F. de Bernardis, R. Bean and M. D. Niemack, *Constraints on massive neutrinos from the pairwise kinematic Sunyaev-Zel'dovich effect*, *Phys. Rev. D* **92** (Sept., 2015) 063501, [1412.0592].

[110] J. Goodman, *Geocentrism reexamined*, *Phys. Rev. D* **52** (Aug., 1995) 1821–1827, [astro-ph/9506068].

[111] J. P. Zibin and A. Moss, *Nowhere to hide: closing in on cosmological homogeneity*, 1409.3831.

[112] U.-L. Pen and P. Zhang, *Observational consequences of dark energy decay*, *Phys. Rev. D* **89** (Mar., 2014) 063009, [1202.0107].

[113] P. Zhang and M. C. Johnson, *Testing eternal inflation with the kinetic Sunyaev Zel'dovich effect*, *JCAP* **1506** (2015) 046, [1501.00511].

[114] P. Van Nieuwenhuizen, *On ghost-free tensor lagrangians and linearized gravitation*, *Nucl. Phys.* **B60** (1973) 478–492.

[115] H. van Dam and M. Veltman, *Massive and massless Yang-Mills and gravitational fields*, *Nucl.Phys.* **B22** (1970) 397–411.

[116] Particle Data Group collaboration, J. Beringer, J. F. Arguin, R. M. Barnett, K. Copic, O. Dahl, D. E. Groom et al., *Review of particle physics**, *Phys. Rev. D* **86** (Jul, 2012) 010001.

[117] E. Babichev and C. Deffayet, *An introduction to the Vainshtein mechanism*, *Class.Quant.Grav.* **30** (2013) 184001, [1304.7240].

[118] T. Damour, I. I. Kogan and A. Papazoglou, *Nonlinear bigravity and cosmic acceleration*, *Phys. Rev.* **D66** (2002) 104025, [hep-th/0206044].

[119] D. Boulware and S. Deser, *Can gravitation have a finite range?*, *Phys.Rev.* **D6** (1972) 3368–3382.

[120] R. Arnowitt, S. Deser and C. W. Misner, *Republication of: The dynamics of general relativity*, *General Relativity and Gravitation* **40** (Sept., 2008) 1997–2027, [gr-qc/0405109].

[121] P. Creminelli, A. Nicolis, M. Papucci and E. Trincherini, *Ghosts in massive gravity*, *Journal of High Energy Physics* **9** (Sept., 2005) 003, [hep-th/0505147].

[122] S. Hassan and R. A. Rosen, *Resolving the Ghost Problem in non-Linear Massive Gravity*, *Phys.Rev.Lett.* **108** (2012) 041101, [1106.3344].

[123] N. Arkani-Hamed, H. Georgi and M. D. Schwartz, *Effective field theory for massive gravitons and gravity in theory space*, *Annals Phys.* **305** (2003) 96–118, [hep-th/0210184].

[124] C. de Rham, G. Gabadadze, L. Heisenberg and D. Pirtskhalava, *Nonrenormalization and naturalness in a class of scalar-tensor theories*, *Phys. Rev. D* **87** (Apr., 2013) 085017, [1212.4128].

[125] A. Nicolis, R. Rattazzi and E. Trincherini, *Galileon as a local modification of gravity*, *Phys. Rev. D* **79** (Mar., 2009) 064036, [0811.2197].

[126] C. de Rham, G. Gabadadze, L. Heisenberg and D. Pirtskhalava, *Cosmic acceleration and the helicity-0 graviton*, *Phys. Rev. D* **83** (May, 2011) 103516, [1010.1780].

[127] R. Woodard, *Avoiding Dark Energy with 1/R Modifications of Gravity*, in *The Invisible Universe: Dark Matter and Dark Energy* (L. Papantonopoulos, ed.), vol. 720 of *Lecture Notes in Physics, Berlin Springer Verlag*, p. 403, 2007. astro-ph/0601672.

[128] M. Luty, M. Porrati and R. Rattazzi, *Strong Interactions and Stability in the DGP Model*, *JHEP* **029** (2003) , [hep-th/0303116].

[129] A. Nicolis and R. Rattazzi, *Classical and Quantum Consistency of the DGP Model*, *JHEP* **6** (2004) 59, [hep-th/0404159].

[130] K. Hinterbichler, M. Trodden and D. Wesley, *Multifield Galileons and higher codimension branes*, *Phys. Rev. D* **82** (Dec., 2010) 124018, [1008.1305].

[131] P. de Fromont, C. de Rham, L. Heisenberg and A. Matas, *Superluminality in the Bi- and Multi- Galileon*, *ArXiv e-prints* (Mar., 2013) , [1303.0274].

[132] C. Burrage, C. de Rham, L. Heisenberg and A. J. Tolley, *Chronology protection in Galileon models and massive gravity*, *JCAP* **7** (July, 2012) 004, [1111.5549].

[133] C. de Rham, M. Fasiello and A. J. Tolley, *Galileon duality*, *Physics Letters B* **733** (June, 2014) 46–51, [1308.2702].

[134] C. de Rham, L. Keltner and A. J. Tolley, *Generalized Galileon duality*, *Phys. Rev. D* **90** (July, 2014) 024050, [1403.3690].

[135] K. Hinterbichler and A. Joyce, *Hidden symmetry of the Galileon*, *Phys. Rev. D* **92** (July, 2015) 023503, [1501.07600].

[136] J. Noller, V. Sivanesan and M. von Strauss, *On extended symmetries for the Galileon*, *Phys. Rev. D* **92** (Sept., 2015) 064009, [1506.03446].

[137] G. Dvali, G. Gabadadze and M. Porrati, *4-D gravity on a brane in 5-D Minkowski space*, *Phys.Lett.* **B485** (2000) 208–214, [hep-th/0005016].

[138] C. Deffayet, *Cosmology on a brane in minkowski bulk*, *Physics Letters B* **502** (2001) 199 – 208.

[139] A. Vainshtein, *To the problem of nonvanishing gravitation mass*, *Phys.Lett.* **B39** (1972) 393–394.

[140] C. Deffayet, G. R. Dvali, G. Gabadadze and A. I. Vainshtein, *Nonperturbative continuity in graviton mass versus perturbative discontinuity*, *Phys. Rev.* **D65** (2002) 044026, [hep-th/0106001].

[141] G. W. Gibbons and S. W. Hawking, *Action integrals and partition functions in quantum gravity*, *Phys. Rev. D* **15** (May, 1977) 2752–2756.

[142] K. Koyama, *Are there ghosts in the self-accelerating brane universe?*, *Phys.Rev.* **D72** (2005) 123511, [hep-th/0503191].

[143] D. Gorbunov, K. Koyama and S. Sibiryakov, *More on ghosts in DGP model*, *Phys.Rev.* **D73** (2006) 044016, [hep-th/0512097].

[144] C. Charmousis, R. Gregory, N. Kaloper and A. Padilla, *DGP Specteroscopy*, *JHEP* **0610** (2006) 066, [hep-th/0604086].

[145] A. De Felice, A. Emir Gümrükçüoğlu, C. Lin and S. Mukohyama, *On the cosmology of massive gravity*, *Classical and Quantum Gravity* **30** (Sept., 2013) 184004, [1304.0484].

[146] G. D'Amico, C. de Rham, S. Dubovsky, G. Gabadadze, D. Pirtskhalava and A. J. Tolley, *Massive cosmologies*, *Phys. Rev. D* **84** (Dec., 2011) 124046, [1108.5231].

[147] A. Emir Gümrükçüoğlu, C. Lin and S. Mukohyama, *Open FRW universes and self-acceleration from nonlinear massive gravity*, *JCAP* **11** (Nov., 2011) 030, [1109.3845].

[148] K. Koyama, G. Niz and G. Tasinato, *Analytic Solutions in Nonlinear Massive Gravity*, *Physical Review Letters* **107** (Sept., 2011) 131101, [1103.4708].

[149] L. Berezhiani, G. Chkareuli, C. de Rham, G. Gabadadze and A. J. Tolley, *On black holes in massive gravity*, *Phys. Rev. D* **85** (Feb., 2012) 044024, [1111.3613].

[150] A. De Felice, A. E. Gümrükçüoğlu and S. Mukohyama, *Massive Gravity: Nonlinear Instability of a Homogeneous and Isotropic Universe*, *Physical Review Letters* **109** (Oct., 2012) 171101, [1206.2080].

[151] A. E. Gümrükçüoğlu, C. Lin and S. Mukohyama, *Anisotropic Friedmann-Robertson-Walker universe from nonlinear massive gravity*, *Physics Letters B* **717** (Oct., 2012) 295–298, [1206.2723].

[152] A. De Felice, A. Emir Gümrükçüoğlu, C. Lin and S. Mukohyama, *Nonlinear stability of cosmological solutions in massive gravity*, *JCAP* **5** (May, 2013) 035, [1303.4154].

[153] S. Hassan and R. A. Rosen, *Bimetric Gravity from Ghost-free Massive Gravity*, *JHEP* **1202** (2012) 126, [1109.3515].

[154] S. Hassan and R. A. Rosen, *Confirmation of the Secondary Constraint and Absence of Ghost in Massive Gravity and Bimetric Gravity*, *JHEP* **1204** (2012) 123, [1111.2070].

[155] C. de Rham, L. Heisenberg and R. H. Ribeiro, *On couplings to matter in massive (bi-)gravity*, *ArXiv e-prints* (Aug., 2014) , [1408.1678].

[156] C. de Rham and A. J. Tolley, *Vielbein to the Rescue?*, *ArXiv e-prints* (May, 2015) , [1505.01450].

[157] S. F. Hassan and R. A. Rosen, *Bimetric gravity from ghost-free massive gravity*, *Journal of High Energy Physics* **2** (Feb., 2012) 126, [1109.3515].

[158] A. R. Solomon, Y. Akrami and T. S. Koivisto, *Linear growth of structure in massive bigravity*, *ArXiv e-prints* (Apr., 2014) , [1404.4061].

[159] F. Könnig, A. Patil and L. Amendola, *Viable cosmological solutions in massive bimetric gravity*, *jcap* **3** (Mar., 2014) 29, [1312.3208].

[160] M. Berg, I. Buchberger, J. Enander, E. Mörtsell and S. Sjörs, *Growth histories in bimetric massive gravity*, *jcap* **12** (Dec., 2012) 21, [1206.3496].

[161] Y. Akrami, T. S. Koivisto and M. Sandstad, *Cosmological constraints on ghost-free bigravity: background dynamics and late-time acceleration*, *ArXiv e-prints* (Feb., 2013) , [1302.5268].

[162] M. von Strauss, A. Schmidt-May, J. Enander, E. Mörtsell and S. F. Hassan, *Cosmological solutions in bimetric gravity and their observational tests*, *jcap* **3** (Mar., 2012) 42, [1111.1655].

[163] M. Lagos and P. G. Ferreira, *Cosmological perturbations in massive bigravity*, *ArXiv e-prints* (Sept., 2014) , [1410.0207].

[164] F. Könnig and L. Amendola, *A minimal bimetric gravity model that fits cosmological observations*, *ArXiv e-prints* (Feb., 2014) , [1402.1988v1].

[165] D. Comelli, M. Crisostomi and L. Pilo, *Perturbations in massive gravity cosmology*, *Journal of High Energy Physics* **6** (June, 2012) 85, [1202.1986].

[166] F. Könnig and L. Amendola, *Instability in a minimal bimetric gravity model*, *Phys. Rev. D* **90** (Aug., 2014) 044030, [1402.1988v2].

[167] F. Könnig, Y. Akrami, L. Amendola, M. Motta and A. R. Solomon, *Stable and unstable cosmological models in bimetric massive gravity*, *ArXiv e-prints* (July, 2014) , [1407.4331].

[168] P. Gratia, W. Hu and M. Wyman, *Self-accelerating massive gravity: how zweibeins walk through determinant singularities*, *Classical and Quantum Gravity* **30** (Sept., 2013) 184007, [1305.2916].

[169] P. Gratia, W. Hu and M. Wyman, *Self-accelerating massive gravity: Bimetric determinant singularities*, *Phys. Rev. D* **89** (Jan., 2014) 027502, [1309.5947].

[170] V. Zakharov, *Linearized gravitation theory and the graviton mass*, *JETP Lett.* **12** (1970) 312.

[171] E. Babichev, C. Deffayet and R. Ziour, *The Vainshtein mechanism in the decoupling limit of massive gravity*, *JHEP* **5** (May, 2009) 98, [0901.0393].

[172] E. Babichev, C. Deffayet and R. Ziour, *Recovering General Relativity from massive gravity*, *Phys.Rev.Lett.* **103** (2009) 201102, [0907.4103].

[173] E. Babichev, C. Deffayet and R. Ziour, *The Recovery of General Relativity in massive gravity via the Vainshtein mechanism*, *Phys.Rev.* **D82** (2010) 104008, [1007.4506].

[174] G. Chkareuli and D. Pirtskhalava, *Vainshtein Mechanism In Λ_3 - Theories*, *Phys.Lett.* **B713** (2012) 99–103, [1105.1783].

[175] F. Sbisa, G. Niz, K. Koyama and G. Tasinato, *Characterising Vainshtein Solutions in Massive Gravity*, *Phys.Rev.* **D86** (2012) 024033, [1204.1193].

[176] L. Berezhiani, G. Chkareuli, C. de Rham, G. Gabadadze and A. Tolley, *Mixed Galileons and Spherically Symmetric Solutions*, *Class.Quant.Grav.* **30** (2013) 184003, [1305.0271].

[177] L. Berezhiani, G. Chkareuli and G. Gabadadze, *Restricted Galileons*, *Phys.Rev.* **D88** (2013) 124020, [1302.0549].

[178] E. Babichev and M. Crisostomi, *Restoring general relativity in massive bigravity theory*, *Phys.Rev.* **D88** (2013) 084002, [1307.3640].

[179] C. de Rham, A. J. Tolley and D. H. Wesley, *Vainshtein Mechanism in Binary Pulsars*, 1208.0580.

[180] C. de Rham, A. Matas and A. J. Tolley, *Galileon Radiation from Binary Systems*, 1212.5212.

[181] Y.-Z. Chu and M. Trodden, *Retarded Green's Function Of A Vainshtein System And Galileon Waves*, *Phys.Rev.* **D87** (2013) 024011, [1210.6651].

[182] E. Babichev, *Plane waves in the generalized Galileon theory*, *Phys.Rev.* **D86** (2012) 084037, [1207.4764].

[183] S. Deser and A. Waldron, *Acausality of Massive Gravity*, *Phys.Rev.Lett.* **110** (2013) 111101, [1212.5835].

[184] S. Deser, K. Izumi, Y. Ong and A. Waldron, *Massive Gravity Acausality Redux*, *Phys.Lett.* **B726** (2013) 544–548, [1306.5457].

[185] S. Deser, K. Izumi, Y. Ong and A. Waldron, *Superluminal Propagation and Acausality of Nonlinear Massive Gravity*, 1312.1115.

[186] S. Deser, M. Sandora, A. Waldron and G. Zahariade, *Covariant constraints for generic massive gravity and analysis of its characteristics*, 1408.0561.

[187] J. Chagoya, K. Koyama, G. Niz and G. Tasinato, *Galileons and strong gravity*, 1407.7744.

[188] C. D. Leonard, J. Ziprick, G. Kunstatter and R. B. Mann, *Gravitational collapse of K-essence Matter in Painlevé-Gullstrand coordinates*, *JHEP* **1110** (2011) 028, [1106.2054].

[189] K. Koyama, G. Niz and G. Tasinato, *Effective theory for the Vainshtein mechanism from the Horndeski action*, *Phys.Rev.* **D88** (2013) 021502, [1305.0279].

[190] A. L. Erickcek, N. Barnaby, C. Burrage and Z. Huang, *Chameleons in the Early Universe: Kicks, Rebounds, and Particle Production*, *Phys.Rev.* **D89** (2014) 084074, [1310.5149].

[191] L. Berezhiani, G. Chkareuli and G. Gabadadze, *Restricted Galileons*, *Phys. Rev. D* **88** (Dec., 2013) 124020, [1302.0549].

[192] K. Hinterbichler, A. Nicolis and M. Porrati, *Superluminality in DGP*, *Journal of High Energy Physics* **9** (Sept., 2009) 89, [0905.2359].

[193] R. Gregory, *The Three burials of Melquiades DGP*, *Prog.Theor.Phys.Suppl.* **172** (2008) 71–80, [0801.1603].

[194] E. Berti, V. Cardoso and A. O. Starinets, *Quasinormal modes of black holes and black branes*, *Class.Quant.Grav.* **26** (2009) 163001, [0905.2975].

[195] E. Barausse, V. Cardoso and P. Pani, *Can environmental effects spoil precision gravitational-wave astrophysics?*, *Phys.Rev.* **D89** (2014) 104059, [1404.7149].

[196] E. Ching, P. Leung, W. Suen and K. Young, *Late time tail of wave propagation on curved space-time*, *Phys.Rev.Lett.* **74** (1995) 2414–2417, [gr-qc/9410044].

[197] E. Ching, P. Leung, W. Suen and K. Young, *Wave propagation in gravitational systems: Late time behavior*, *Phys.Rev.* **D52** (1995) 2118–2132, [gr-qc/9507035].

[198] K. D. Kokkotas and B. G. Schmidt, *Quasinormal modes of stars and black holes*, *Living Rev.Rel.* **2** (1999) 2, [gr-qc/9909058].

[199] V. Cardoso, A. S. Miranda, E. Berti, H. Witek and V. T. Zanchin, *Geodesic stability, Lyapunov exponents and quasinormal modes*, *Phys.Rev.* **D79** (2009) 064016, [0812.1806].

[200] R. Konoplya and A. Zhidenko, *Quasinormal modes of black holes: From astrophysics to string theory*, *Rev.Mod.Phys.* **83** (2011) 793–836, [1102.4014].

[201] S. Chandrasekhar and S. L. Detweiler, *The quasi-normal modes of the Schwarzschild black hole*, *Proc.Roy.Soc.Lond.* **A344** (1975) 441–452.

[202] P. Pani, *Advanced Methods in Black-Hole Perturbation Theory*, *Int.J.Mod.Phys.* **A28** (2013) 1340018, [1305.6759].

[203] R. Akhoury, D. Garfinkle and R. Saotome, *Gravitational collapse of k-essence*, *JHEP* **1104** (2011) 096, [1103.0290].

[204] R. Akhoury, D. Garfinkle, R. Saotome and A. Vikman, *Non-Stationary Dark Energy Around a Black Hole*, *Phys.Rev.* **D83** (2011) 084034, [1103.2454].

[205] E. Babichev, V. F. Mukhanov and A. Vikman, *Escaping from the black hole?*, *JHEP* **0609** (2006) 061, [hep-th/0604075].

[206] E. Babichev, *Galileon accretion*, *Phys.Rev.* **D83** (2011) 024008, [1009.2921].

[207] C. Burrage, C. de Rham, L. Heisenberg and A. J. Tolley, *Chronology Protection in Galileon Models and Massive Gravity*, *JCAP* **1207** (2012) 004, [1111.5549].

[208] E. Babichev, V. Mukhanov and A. Vikman, *k-Essence, superluminal propagation, causality and emergent geometry*, *JHEP* **0802** (2008) 101, [0708.0561].

[209] K. Izumi, K. Koyama, O. Pujolas and T. Tanaka, *Bubbles in the Self-Accelerating Universe*, *Phys.Rev.* **D76** (2007) 104041, [0706.1980].

[210] M. Park and L. Sorbo, *Vacua and instantons of ghost-free massive gravity*, *Phys.Rev.* **D87** (2013) 024041, [1212.2691].

[211] J. Sakstein, *Disformal Theories of Gravity: From the Solar System to Cosmology*, 1409.1734.

[212] P. Brax and C. Burrage, *Constraining Disformally Coupled Scalar Fields*, 1407.1861.

[213] G. Dvali, G. F. Giudice, C. Gomez and A. Kehagias, *UV-Completion by Classicalization*, *JHEP* **1108** (2011) 108, [1010.1415].

[214] G. Dvali, C. Gomez and A. Kehagias, *Classicalization of Gravitons and Goldstones*, *JHEP* **1111** (2011) 070, [1103.5963].

[215] G. Dvali and D. Pirtskhalava, *Dynamics of Unitarization by Classicalization*, *Phys.Lett.* **B699** (2011) 78–86, [1011.0114].

[216] G. Cusin, R. Durrer, P. Guarato and M. Motta, *Gravitational waves in bigravity cosmology*, *ArXiv e-prints* (Dec., 2014) , [1412.5979].

[217] L. Amendola, F. Könnig, M. Martinelli, V. Pettorino and M. Zumalacarregui, *Surfing gravitational waves: can bigravity survive growing tensor modes?*, 1503.02490.

[218] A. De Felice, T. Nakamura and T. Tanaka, *Possible existence of viable models of bi-gravity with detectable graviton oscillations by gravitational wave detectors*, *PTEP* **2014** (2014) 043E01, [1304.3920].

[219] R. Brito, V. Cardoso and P. Pani, *Massive spin-2 fields on black hole spacetimes: Instability of the Schwarzschild and Kerr solutions and bounds on the graviton mass*, *Phys.Rev.* **D88** (2013) 023514, [1304.6725].

[220] I. D. Saltas, I. Sawicki, L. Amendola and M. Kunz, *Anisotropic Stress as a Signature of Nonstandard Propagation of Gravitational Waves*, *Phys.Rev.Lett.* **113** (2014) 191101, [1406.7139].

[221] M. Raveri, C. Baccigalupi, A. Silvestri and S.-Y. Zhou, *Measuring the speed of cosmological gravitational waves*, 1405.7974.

[222] L. Xu, *Gravitational Waves: A Test for Modified Gravity*, 1410.6977.

[223] V. Pettorino and L. Amendola, *Friction in Gravitational Waves: a test for early-time modified gravity*, *Phys.Lett.* **B742** (2015) 353–357, [1408.2224].

[224] Y. Cai, Y.-T. Wang and Y.-S. Piao, *Oscillating modulation to B-mode polarization from varying propagating speed of primordial gravitational waves*, 1501.06345.

[225] B. Abbott, R. Abbott, R. Adhikari, J. Agresti, P. Ajith, B. Allen et al., *Searching for a Stochastic Background of Gravitational Waves with the Laser Interferometer Gravitational-Wave Observatory*, *Astrophysical Journal* **659** (Apr., 2007) 918–930, [astro-ph/0608606].

[226] F. A. Jenet, G. B. Hobbs, W. van Straten, R. N. Manchester, M. Bailes, J. P. W. Verbiest et al., *Upper Bounds on the Low-Frequency Stochastic Gravitational Wave Background from Pulsar Timing Observations: Current Limits and Future Prospects*, *Astrophysical Journal* **653** (Dec., 2006) 1571–1576, [astro-ph/0609013].

[227] C. Ungarelli and A. Vecchio, *Studying the anisotropy of the gravitational wave stochastic background with LISA*, *Phys. Rev. D* **64** (Dec., 2001) 121501, [astro-ph/0106538].

[228] C. Cutler and J. Harms, *Big Bang Observer and the neutron-star-binary subtraction problem*, *Phys. Rev. D* **73** (Feb., 2006) 042001, [gr-qc/0511092].

[229] L. A. Boyle and P. J. Steinhardt, *Probing the early universe with inflationary gravitational waves*, *Phys. Rev. D* **77** (Mar., 2008) 063504, [astro-ph/0512014].

[230] S. Kuroyanagi, S. Tsujikawa, T. Chiba and N. Sugiyama, *Implications of the B-mode polarization measurement for direct detection of inflationary gravitational waves*, *Phys. Rev. D* **90** (Sept., 2014) 063513, [1406.1369].

[231] T. S. Bunch and P. C. W. Davies, *Quantum field theory in de sitter space: Renormalization by point-splitting*, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **360** (1978) 117–134, [http://rspa.royalsocietypublishing.org/content/360/1700/117.full.pdf].

[232] DESI collaboration, A. Aghamousa et al., *The DESI Experiment Part I: Science,Targeting, and Survey Design*, 1611.00036.

[233] LSST Science Collaboration, P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel et al., *LSST Science Book, Version 2.0*, *ArXiv e-prints* (Dec., 2009) , [0912.0201].

[234] K. Bandura, G. E. Addison, M. Amiri, J. R. Bond, D. Campbell-Wilson, L. Connor et al., *Canadian Hydrogen Intensity Mapping Experiment (CHIME) pathfinder*, in *Ground-based and Airborne Telescopes V*, vol. 9145, p. 914522, July, 2014. 1406.2288. DOI.

[235] Cosmology SWG collaboration, F. B. Abdalla et al., *Cosmology from HI galaxy surveys with the SKA*, 1501.04035.

[236] J. R. Pritchard and A. Loeb, *21 cm cosmology in the 21st century*, *Reports on Progress in Physics* **75** (2012) 086901, [1109.6012].

[237] A. L. Erickcek, S. M. Carroll and M. Kamionkowski, *Superhorizon perturbations and the cosmic microwave background*, *Phys. Rev. D* **78** (Oct., 2008) 083012, [0808.1570].

[238] E. Bertschinger, *Multiscale gaussian random fields and their application to cosmological simulations*, *The Astrophysical Journal Supplement Series* **137** (2001) 1.

[239] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, *Astrophysical Journal* **622** (Apr., 2005) 759–771, [astro-ph/0409513].

[240] VIRGO Consortium collaboration, R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce et al., *Stable clustering, the halo model and nonlinear cosmological power spectra*, *Mon. Not. Roy. Astron. Soc.* **341** (2003) 1311, [astro-ph/0207664].

[241] A. R. Pullen and M. Kamionkowski, *Cosmic Microwave Background Statistics for a Direction-Dependent Primordial Power Spectrum*, *Phys. Rev.* **D76** (2007) 103529, [0709.1144].

[242] M. Loverde and N. Afshordi, *Extended Limber approximation*, *Phys. Rev. D* **78** (Dec., 2008) 123506, [0809.5112].

[243] C.-P. Ma and J. N. Fry, *Nonlinear Kinetic Sunyaev-Zeldovich Effect*, *Physical Review Letters* **88** (May, 2002) 211301, [astro-ph/0106342].

[244] H. Park, E. Komatsu, P. R. Shapiro, J. Koda and Y. Mao, *The Impact of Nonlinear Structure Formation on the Power Spectrum of Transverse Momentum Fluctuations and the Kinetic Sunyaev-Zel'dovich Effect*, *Astrophysical Journal* **818** (Feb., 2016) 37.

[245] W. Hu, *Reionization Revisited: Secondary Cosmic Microwave Background Anisotropies and Polarization*, *Astrophysical Journal* **529** (Jan., 2000) 12–25, [astro-ph/9907103].

[246] CMB-S4 collaboration, K. N. Abazajian et al., *CMB-S4 Science Book, First Edition*, 1610.02743.

[247] R. Laureijs, J. Amiaux, S. Arduini, J. . Augueres, J. Brinchmann, R. Cole et al., *Euclid Definition Study Report*, *ArXiv e-prints* (Oct., 2011) , [1110.3193].

[248] H. C. Ferguson, *The Hubble Deep Field*, in *Reviews in Modern Astronomy* (R. E. Schielicke, ed.), vol. 11 of *Reviews in Modern Astronomy*, p. 83, 1998.

[249] S. V. W. Beckwith et al., *The Hubble Ultra Deep Field*, *Astron. J.* **132** (2006) 1729–1755, [astro-ph/0607632].

[250] Planck collaboration, N. Aghanim et al., *Planck 2013 results. XXVII. Doppler boosting of the CMB: Eppur si muove*, *Astron. Astrophys.* **571** (2014) A27, [1303.5087].

[251] C. Gibelyou and D. Huterer, *Dipoles in the sky*, *Monthly Notices of the Royal Astronomical Society* **427** (2012) 1994–2021, [http://mnras.oxfordjournals.org/content/427/3/1994.full.pdf+html].

[252] M. Yoon and D. Huterer, *Kinematic dipole detection with galaxy surveys: forecasts and requirements*, *Astrophys. J.* **813** (2015) L18, [1509.05374].

[253] N. Kaiser, *On the spatial correlations of Abell clusters*, *The Astrophysical Journal, Letters* **284** (Sept., 1984) L9–L12.

[254] D. C. Fox and A. Loeb, *Do the electrons and ions in X-ray clusters share the same temperature?*, *Astrophys. J.* **491** (1997) 459, [astro-ph/9706266].

[255] M. Zaldarriaga, *Lensing of the CMB: Non-Gaussian aspects*, *Phys. Rev.* **D62** (2000) 063510, [astro-ph/9910498].

[256] N. Kaiser, *Clustering in real space and in redshift space*, *MNRAS* **227** (July, 1987) 1–21.

[257] J. Delabrouille, J.-L. Puget, J.-M. Lamarre and R. Gispert, *Scanning Strategies for the Planck Mission*, *Astrophysical Letters and Communications* **37** (2000) 259.

[258] S. Adhikari, S. Shandera and A. L. Erickcek, *Large-scale anomalies in the cosmic microwave background as signatures of non-Gaussianity*, *Phys. Rev.* **D93** (2016) 023524, [1508.06489].

[259] A. P. S. Yadav and B. D. Wandelt, *Cmb tomography: Reconstruction of adiabatic primordial scalar potential using temperature and polarization maps*, *Physical Review D* **71** (jun, 2005) 123004, [0505386].

[260] K. Koyama, *Cosmological Tests of Modified Gravity*, *Rept. Prog. Phys.* **79** (2016) 046902, [1504.04623].

[261] B. Jain, A. Joyce, R. Thompson, A. Upadhye, J. Battat et al., *Novel Probes of Gravity and Dark Energy*, 1309.5389.

[262] D. Huterer, D. Kirkby, R. Bean, A. Connolly, K. Dawson et al., *Growth of Cosmic Structure: Probing Dark Energy Beyond Expansion*, 1309.5385.

[263] Euclid Theory Working Group collaboration, L. Amendola et al., *Cosmology and fundamental physics with the Euclid satellite*, *Living Rev. Rel.* **16** (2013) 6, [1206.1225].

[264] R. Courant and D. Hibert, *Methods of Mathematical Physics: Partial Differential Equations*, vol. II. Wiley, 1961.

Appendix A

Convergence properties

The convergence properties of our numerical results can be understood by varying the grid size. For a p^{th} order scheme, the convergence ratio defined by

$$Q = \frac{||\pi^{4h} - \pi^{2h}||_2}{||\pi^{2h} - \pi^{h}||_2} \tag{A.1}$$

yields $Q = 2^p$ in the continuum limit, $h \to 0$. Here, the superscript on the numerical solution π refers to the size of the spacing of the grid used, and $||\cdot||_2$ is the ℓ_2-norm. For example, for the second order scheme we use, $Q \simeq 4$. We have evolved both sets of initial conditions introduced in Sections 3.3.2–3.3.2, in particular eq. (3.16) with $\rho = 200$, $R_0 = 1$, $A = 0.002$, $\sigma = 1$, $r_w = 10$, $\epsilon = 0.001$ and $\pi(r, 0) = 0$, respectively. The results are summarized in figure A.1, and are compatible with second-order convergence.

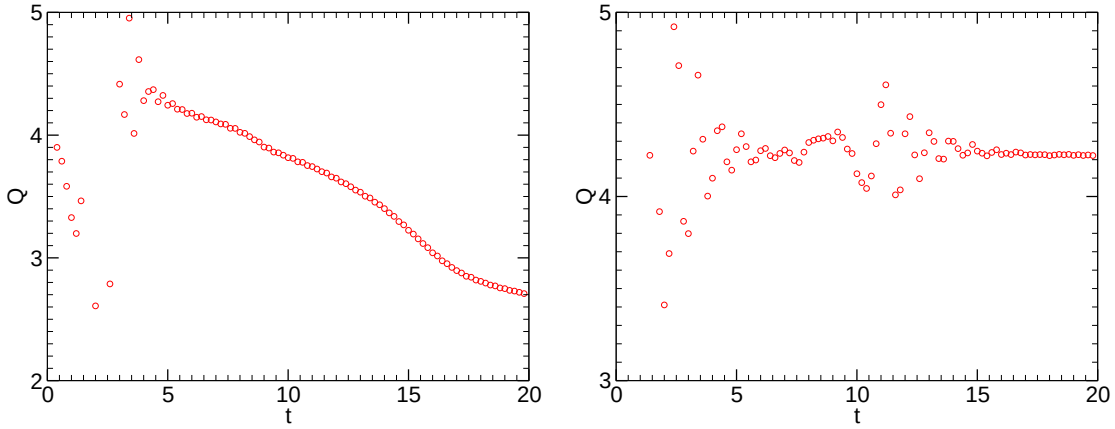

Figure A.1: The convergence factor $Q(t)$ defined by eq. (A.1) as a function of time. Left panel refers to initial data of the form $\pi(r, 0) = 0$, $\rho = 50$, $R_0 = 1$, while the right panel refers to eq. (3.16) for $\rho = 200$, $R_0 = 1$, $A = 0.002$, $\sigma = 1$, $r_w = 10$, $\epsilon = 0.001$.

Appendix B

Cauchy problem and stability

Let us first consider the Lagrangian (3.2) for a spherically symmetric field $\pi(t,r)$. The equations of motion coming from this Lagrangian have the form of a Monge-Ampére equation

$$A\ddot{\pi} + B\dot{\pi}' + C\pi'' + D + E\left[\ddot{\pi}\pi'' - (\dot{\pi}')^2\right] , \tag{B.1}$$

where A, B, C and D are at most functions of π and its first derivatives, and we assume $(A + E\pi'') \neq 0$. We wish to understand when this type of equation plus its initial conditions describe a well-posed initial value problem, commonly known as Cauchy problem (see e.g. Chapter V of [264] and Appendix 1 where the Cauchy problem for the Monge-Ampére equation is considered). A family of curves $\varphi(t,r) = 0$ are characteristics of this equation if

$$(A + E\pi'')\varphi_t^2 + (B - 2E\dot{\pi}')\varphi_t\varphi_r + (C + E\ddot{\pi})\varphi_r^2 = 0\,, \tag{B.2}$$

where $\varphi_r \equiv \partial_r\varphi$ and $\varphi_t \equiv \partial_t\varphi$. Along a characteristic curve solutions are constant, so we can write

$$\frac{d\varphi}{d\lambda} = \varphi_r\frac{dr}{d\lambda} + \varphi_t\frac{dt}{d\lambda} = 0\,, \tag{B.3}$$

where λ denotes the parameter along which the characteristic curve is constant. Substituting in eq. (B.2), we have

$$(A + E\pi'')\left(\frac{dr/d\lambda}{dt/d\lambda}\right)^2 - (B - 2E\dot{\pi}')\frac{dr/d\lambda}{dt/d\lambda} + (C + E\ddot{\pi}) = 0\,. \tag{B.4}$$

Solving with respect to $\frac{dr/d\lambda}{dt/d\lambda} \equiv dr/dt$ we find two roots

$$u_1 \equiv \frac{dr}{dt} = \frac{B - 2E\tilde{s} + \Delta}{2(A + E\tilde{r})}\,, \tag{B.5}$$

$$u_2 \equiv \frac{dr}{dt} = \frac{B - 2E\tilde{s} - \Delta}{2(A + E\tilde{r})}\,, \tag{B.6}$$

where $\Delta^2 = B^2 - 4AC + 4DE$ and we used eq. (B.1) in the form

$$(A + E\pi'')(C + E\ddot{\pi}) - \frac{1}{4}(B - 2E\dot{\pi}')^2 + \frac{1}{4}\Delta^2 = 0\,. \tag{B.7}$$

The type of this equation is determined by the discriminant Δ^2:

- If $\Delta^2 > 0$, the equation is hyperbolic (two roots);
- If $\Delta^2 = 0$, the equation is parabolic (one root);
- If $\Delta^2 < 0$, the equation is elliptic (imaginary roots).

For galileon-like models, the discriminant Δ depends on the first derivatives of π. A well-posed initial value problem is equivalent to requiring that the equation be hyperbolic, i.e. $\Delta^2 > 0$. Note that we also must require the initial conditions to satisfy $A + E\pi'' < 0$, which is equivalent to requiring the constant time $t = t_0$ hypersurface to be spacelike everywhere.

Let us see how this can be understood in terms of scalar perturbations $\delta\pi$ about a background π_0. The action for $\delta\pi$ is given by

$$S_{\delta\pi} = \int d^4x \left[\frac{1}{2} Z^{\mu\nu} \partial_\mu \delta\pi \partial_\nu \delta\pi\right], \tag{B.8}$$

where $Z^{\mu\nu}$ is the effective metric, dependent on π_0 and its derivatives, on which $\delta\pi$ propagates. Requiring the equations of motion for $\delta\pi$ to be hyperbolic is equivalent to requiring the effective metric $Z^{\mu\nu}$ to have a Lorentzian signature, i.e., $\det(Z^{\mu\nu}) < 0$, which in a spherically symmetric spacetime can be written as

$$\det(Z^{\mu\nu}) = [Z^{tt} Z^{rr} - (Z^{tr})^2](Z^{\theta\theta})^2 \sin^2\theta < 0\,. \tag{B.9}$$

From eq. (B.7) this can be shown to be equivalent to the condition $\Delta^2 > 0$ for the background π_0.

Furthermore, the initial value problem is well posed only if the initial data are set up on a hypersurface Σ which is spacelike with respect to the effective metric $Z^{-1}_{\mu\nu}$, where $Z^{-1}_{\mu\nu}$ is defined such that $Z^{\mu\lambda} Z^{-1}_{\lambda\nu} = \delta^\mu_\nu$, i.e., we require the 1-form $\partial_\mu t$ to be timelike with respect to $Z^{\mu\nu}$, $Z^{\mu\nu}\partial_\mu t \partial_\mu t < 0$. Thus the initial data must be such that $Z^{tt} < 0$, which is equivalent to requiring $A + E\pi'' < 0$ for the background π_0. However, note that in general even for an initially well-posed Cauchy problem, due to the non-linearity of the field equations, the *global* existence and uniqueness of solutions cannot be guaranteed.

As a final remark let us consider the stability under high-frequency perturbations of the background π_0. Local stability at a given point in spacetime p_0 requires the metric to have a Lorentzian signature at p_0 which may appear either in the form of ghost or gradient instabilities [1]. Through this work we have not considered coupling to other dynamical fields, but if this is taken into account we must also require the matrix $Z^{\mu\nu}$ to have the same

[1] Ghost instabilities are characterized either by a wrong-sign of the time-component of the field equations,

signature of the gravitational metric. This requirement is met as long as the matrix $Z^{\mu\nu}$ is non-singular and when diagonalized has the signature $(-,+,+,+)$. In spherical coordinates this is equivalent to requiring

$$\frac{Z^{tt}+Z^{rr}-\sqrt{4(Z^{rt})^2+(Z^{tt}-Z^{rr})^2}}{2}<0\,, \tag{B.10}$$

$$\frac{Z^{tt}+Z^{rr}+\sqrt{4(Z^{rt})^2+(Z^{tt}-Z^{rr})^2}}{2}>0\,, \tag{B.11}$$

$$Z^{\theta\theta}>0\,. \tag{B.12}$$

Note that this requirement is not met in the background of the Π_- branch (see Section 3.4), so this solution would be unstable if it were coupled to another field or if the dynamical degrees of freedom of the source were taken into account, as discussed in the main text.

which is characterized by $\det(Z^{\mu\nu})>0$ when all the other components have the "correct" signature, or by a relative overall difference of sign of the effective metric with respect to the gravitational metric, when coupling to other fields is considered. On the other hand gradient instabilities arise when $\det(Z^{ij})<0$, where the indices i, j take values on the 3-dimensional spacelike hypersurfaces $t=$ const.

Appendix C

The effective velocity in Fourier space

In this appendix, we derive an expression for the effective line of sight velocity $v_{\rm eff}$ in Fourier space. The Fourier transform of the primordial potential is defined as

$$\Psi_i(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3} \tilde{\Psi}_i(\mathbf{k}) e^{i\chi_e \mathbf{k}\cdot\hat{\mathbf{n}}_e} e^{i\Delta\chi \mathbf{k}\cdot\hat{\mathbf{n}}}, \tag{C.1}$$

where we have explicitly expanded the position $\mathbf{r} = \chi_e \hat{\mathbf{n}}_e + \Delta\chi \hat{\mathbf{n}}$.

C.1 Sachs-Wolfe

Using eq. (5.7), the SW contribution to the effective velocity is related to the Fourier components of Ψ_i through

$$\begin{aligned} v_{\rm eff,SW}(\hat{\mathbf{n}}_e, \chi_e) =& \frac{3}{4\pi}\left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2}\right) \int d^2\hat{\mathbf{n}}\ \Psi_i(\mathbf{r})\ \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e) \\ =& \frac{3}{4\pi}\left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2}\right) \int \frac{d^3k}{(2\pi)^3} \tilde{\Psi}_i(\mathbf{k}) e^{i\chi_e \mathbf{k}\cdot\hat{\mathbf{n}}_e} \\ & \times \int d^2\hat{\mathbf{n}}\ e^{i\Delta\chi \mathbf{k}\cdot\hat{\mathbf{n}}}\ \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e). \end{aligned} \tag{C.2}$$

We can work on the second integral by expanding the exponential in terms of Legendre polynomials and spherical Bessel functions:

$$e^{i\Delta\chi \mathbf{k}\cdot\hat{\mathbf{n}}} = \sum_{\ell'} i^{\ell'} (2\ell'+1)\ j_{\ell'}(k\Delta\chi)\ \mathcal{P}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}). \tag{C.3}$$

Substituting and applying the identity,

$$\int d^2\hat{\mathbf{b}}\ \mathcal{P}_{\ell'}(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}})\ \mathcal{P}_\ell(\hat{\mathbf{b}}\cdot\hat{\mathbf{c}}) = \frac{4\pi}{2\ell+1}\mathcal{P}_\ell(\hat{\mathbf{a}}\cdot\hat{\mathbf{c}})\delta_{\ell\ell'}, \tag{C.4}$$

results in

$$\int d^2\hat{\mathbf{n}}\; e^{i\Delta\chi\mathbf{k}\cdot\hat{\mathbf{n}}}\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e) = 4\pi i\; j_1(k\Delta\chi)\; \mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e). \tag{C.5}$$

Putting this result back into eq. (C.2) gives

$$v_{\rm eff,SW}(\hat{\mathbf{n}}_e,\chi_e) = \quad 3i\left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2}\right)\int \frac{d^3k}{(2\pi)^3}\tilde{\Psi}_i(\mathbf{k}) j_1(k\Delta\chi_{\rm dec})\mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e)e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}. \tag{C.6}$$

We therefore see that $v_{\rm eff,SW}$ is simply a convolution of the potential field evaluated on the intersection of the electron's past light cone and the time of decoupling.

C.2 Doppler

Using eq. (5.10), the Doppler contribution to $v_{\rm eff}$ is

$$\begin{aligned} v_{\rm eff,Doppler}(\hat{\mathbf{n}}_e,\chi_e) =& \frac{3}{4\pi}D_v(\chi_{\rm dec})\int d^2\hat{\mathbf{n}}\; (\hat{\mathbf{n}}\cdot\nabla\Psi_i(\mathbf{r}_{\rm dec}))\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e) \\ & - \frac{3}{4\pi}D_v(z_{\rm e})\int d^2\hat{\mathbf{n}}\; (\hat{\mathbf{n}}\cdot\nabla\Psi_i(\mathbf{r}_{\rm e}))\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e). \end{aligned} \tag{C.7}$$

We'll start with the first integral, then do the second. Going to Fourier space and expanding the exponent in Legendre polynomials gives

$$\begin{aligned} & \frac{3}{4\pi}D_v(\chi_{\rm dec})\int \frac{d^3k}{(2\pi)^3} ik\tilde{\Psi}_i(\mathbf{k})\; e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}\int d^2\hat{\mathbf{n}}\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{k}})\; e^{i\Delta\chi_{\rm dec}\mathbf{k}\cdot\hat{\mathbf{n}}}\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e) \\ =& \frac{3}{4\pi}D_v(\chi_{\rm dec})\int \frac{d^3k}{(2\pi)^3} ik\tilde{\Psi}_i(\mathbf{k})\; e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}\sum_{\ell'} i^{\ell'}(2\ell'+1)j_{\ell'}(k\Delta\chi_{\rm dec}) \\ & \times \int d^2\hat{\mathbf{n}}\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{k}})\; \mathcal{P}_{\ell'}(\hat{\mathbf{n}}\cdot\hat{\mathbf{k}})\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e). \end{aligned} \tag{C.8}$$

The integral over three Legendre polynomials can be evaluated by expanding in spherical harmonics and using the Wigner 3j symbols. The result is

$$\int d^2\hat{\mathbf{n}}\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{k}})\; \mathcal{P}_{\ell'}(\hat{\mathbf{n}}\cdot\hat{\mathbf{k}})\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e) = \frac{4\pi}{3}\mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e)\delta_{\ell'0} + \frac{8\pi}{15}\mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e)\delta_{\ell'2}, \tag{C.9}$$

and upon substitution, we obtain for the first integral:

$$iD_v(\chi_{\rm dec})\int \frac{d^3k}{(2\pi)^3}\; k\; \tilde{\Psi}_i(\mathbf{k})\; [(j_0(k\Delta\chi_{\rm dec}) - 2j_2(k\Delta\chi_{\rm dec})]\; \mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e)\; e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}. \tag{C.10}$$

Moving to the second integral in eq. (C.7), we have

$$-\frac{3}{4\pi}D_v(\chi_e)\int \frac{d^3k}{(2\pi)^3} ik\tilde{\Psi}_i(\mathbf{k})e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}\; \int d^2\hat{\mathbf{n}}\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{k}})\; \mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e). \tag{C.11}$$

The integral over angles can be evaluated using the identity eq. (C.4) to obtain

$$-iD_v(\chi_e)\int \frac{d^3k}{(2\pi)^3}\; k\; \tilde{\Psi}_i(\mathbf{k})\; \mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e)\; e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}. \tag{C.12}$$

Assembling the various pieces, the Doppler contribution becomes:

$$v_{\rm eff,Doppler}(\hat{\mathbf{n}}_e,\chi_e) = i\int \frac{d^3k}{(2\pi)^3}\; k\tilde{\Psi}_i(\mathbf{k})\mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e)\; e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e} \\ \times\left[D_v(\chi_{\rm dec})j_0(k\Delta\chi_{\rm dec}) - 2D_v(\chi_{\rm dec})j_2(k\Delta\chi_{\rm dec}) - D_v(\chi_e)\right]. \qquad \text{(C.13)}$$

Note that the last term in square brackets contains no spherical Bessel function. Therefore, the Doppler component receives contributions from all scales, unlike the Sachs-Wolfe term.

C.3 Integrated Sachs-Wolfe

Using eq. (5.14), the ISW contribution to the effective velocity is

$$v_{\rm eff,ISW}(\hat{\mathbf{n}}_e,\chi_e) = \frac{3}{4\pi}\int d^2\hat{\mathbf{n}}\left(2\int_{a_{\rm dec}}^{a_e}\frac{dD_\Psi}{da}\Psi_i(\mathbf{r}(a))da\right)\mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e). \qquad \text{(C.14)}$$

Going to Fourier space,

$$v_{\rm eff,ISW}(\hat{\mathbf{n}}_e,\chi_e) = \frac{3}{2\pi}\int_{a_{\rm dec}}^{a_e} da\frac{dD_\Psi}{da}\int\frac{d^3k}{(2\pi)^3}\tilde{\Psi}_i(\mathbf{k})e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}\int d^2\hat{\mathbf{n}}\; e^{i\Delta\chi(a)\mathbf{k}\cdot\hat{\mathbf{n}}}\mathcal{P}_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}_e), \quad \text{(C.15)}$$

and applying the identity eq. (C.5), we obtain

$$v_{\rm eff,ISW}(\hat{\mathbf{n}}_e,\chi_e) = 6i\int\frac{d^3k}{(2\pi)^3}\tilde{\Psi}_i(\mathbf{k})\left[\int_{a_{\rm dec}}^{a_e} da\frac{dD_\Psi}{da}\; j_1(k\Delta\chi(a))\right]\mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e)e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}. \qquad \text{(C.16)}$$

Just as for the SW contribution, the ISW contribution is mainly sensitive to potential fluctuations on large scales.

C.4 Effective velocity

We can now assemble eqs. (C.6), (C.13), (C.16) into an expression for the total effective velocity. Before doing so, it must be noted that the expression eq. (C.13) for the Doppler kernel is only valid in the small-k limit, so the linear growth with k eventually gets cut off. To fix this, we can incorporate the transfer function $T(k)$ by simply replacing $\tilde{\Psi}_i(\mathbf{k}) \to T(k)\tilde{\Psi}_i(\mathbf{k})$. We will employ the BBKS fitting function:

$$T(k) = \frac{\ln\left[1+0.171x\right]}{0.171x}\left[1+0.284x+(1.18x)^2+(0.399x)^3+(0.49x)^4\right]^{-0.25}, \qquad \text{(C.17)}$$

where $x = k/k_{\rm eq}$ with $k_{\rm eq} = a_{\rm eq}H(a_{\rm eq}) = \sqrt{2/a_{\rm eq}}H_0 \simeq 82.5H_0$. Putting together all three components, (C.6), (C.13), (C.16), gives the expression for the effective velocity:

$$v_{\rm eff}(\hat{\mathbf{n}}_e,\chi_e) = i\int\frac{d^3k}{(2\pi)^3}\; T(k)\tilde{\Psi}_i(\mathbf{k})\;\mathcal{K}^v(k,\chi_e)\;\mathcal{P}_1(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_e)\; e^{i\chi_e\mathbf{k}\cdot\hat{\mathbf{n}}_e}, \qquad \text{(C.18)}$$

where $\mathcal{K}^v = \mathcal{K}_{\rm D} + \mathcal{K}_{\rm SW} + \mathcal{K}_{\rm ISW}$ is the full Fourier kernel with each component given by,

$$\mathcal{K}_{\rm D}(k,\chi_e) \equiv kD_v(\chi_{\rm dec})j_0(k\Delta\chi_{\rm dec}) - 2kD_v(\chi_{\rm dec})j_2(k\Delta\chi_{\rm dec}) - kD_v(\chi_e), \tag{C.19}$$

$$\mathcal{K}_{\rm SW}(k,\chi_e) \equiv 3\left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2}\right) j_1(k\Delta\chi_{\rm dec}), \tag{C.20}$$

$$\mathcal{K}_{\rm ISW}(k,\chi_e) \equiv 6\int_{a_{\rm dec}}^{a_e} da \frac{dD_\Psi}{da}\; j_1(k\Delta\chi(a)). \tag{C.21}$$

Appendix D

A pure gradient is pure gauge

In this appendix we explicitly demonstrate that a pure gradient in the Newtonian potential Ψ (or more generally, the curvature perturbation in an arbitrary gauge) can be removed through a special conformal transformation. More generally, we can remove the gradient of the Newtonian potential at a point, which we take to be the origin of Cartesian coordinates. The Newtonian potential appears in a conformal factor in front of the spatial metric:

$$ds_3^2 = (1 - 2\Psi(\mathbf{x}))\delta_{ij}dx^i dx^j. \tag{D.1}$$

Performing a special conformal transformation

$$x^i = \frac{x'^i - b^i x'_i x'^i}{1 - 2b_i x'^i + (b_i b^i)\left(x'_i x'^i\right)}, \tag{D.2}$$

takes the spatial metric to

$$\delta_{ij}dx^i dx^j = \frac{\delta_{ij}}{\left[1 - 2b_i x'^i + (b_i b^i)\left(x'_i x'^i\right)\right]^2} dx'^i dx'^j, \tag{D.3}$$

where b_i are free constants.

If we imagine there was a pure gradient in the Newtonian potential,

$$\Psi(\mathbf{x}) = A_i x^i, \tag{D.4}$$

we can write

$$\begin{aligned}(1 - 2\Psi(\mathbf{x}))\delta_{ij}dx^i dx^j = &\left(1 - 2A_i \frac{x'^i - b^i x'_i x'^i}{1 - 2b_i x'^i + (b_i b^i)\left(x'_i x'^i\right)}\right) \\ &\times \frac{\delta_{ij}dx'^i dx'^j}{\left[1 - 2b_i x'^i + (b_i b^i)\left(x'_i x'^i\right)\right]^2}.\end{aligned} \tag{D.5}$$

For $b_i \ll 1$, expanding to first order, we have:

$$(1 - 2\Psi(\mathbf{x}))\delta_{ij}dx^i dx^j \simeq \left(1 - 2(A_i - 2b_i)x'^i + \mathcal{O}(b^2)\right)\delta_{ij}dx'^i dx'^j. \tag{D.6}$$

Therefore, with the choice,

$$b_i = A^i/2, \tag{D.7}$$

we have $\Psi(\mathbf{x}') = 0$, and therefore no gradient in the primed coordinate system. Note also that the choice for b_i justifies neglecting the terms of higher order in b, at least in linear perturbation theory.

More generally, we could imagine performing a Taylor series expansion of $\Psi(\mathbf{x})$ about a point, taken here to be $\mathbf{x} = 0$:

$$\Psi(\mathbf{x}) \simeq \Psi(0) + \partial_i\Psi(0)x^i + \dots \tag{D.8}$$

If we perform a special conformal transformation with

$$b_i = \partial_i\Psi(0)/2, \tag{D.9}$$

then in the primed coordinates we have

$$\partial_i\Psi(0)x^i = \frac{1}{2} - \frac{1 - 3\partial_i\Psi(0)\ x'^i + \left(\partial_i\Psi(0)\ \partial^i\Psi(0)\right)\left(x'_i x'^i\right)/4 + (\partial^i\Psi(0)\ x'^i)^2}{2\left[1 - \partial_i\Psi(0)\ x'^i + (\partial_i\Psi(0)\ \partial^i\Psi(0))\left(x'_i x'^i\right)/4\right]^3}. \tag{D.10}$$

To lowest order in $\partial_i\Psi(0)$, this is

$$\partial_i\Psi(0)x^i \simeq (\partial_i\Psi(0)\ x'^i)^2 + \left(\partial_i\Psi(0)\ \partial^i\Psi(0)\right)\left(x'_i x'^i\right)/4. \tag{D.11}$$

Therefore, in the primed coordinate system, the Taylor series expansion of $\Psi(\mathbf{x}')$ is

$$\Psi(\mathbf{x}') \simeq \Psi(0) + \mathcal{O}(x'^2) + \dots \tag{D.12}$$

with no linear term as advertised. Therefore, the special conformal transformation can be used to eliminate the derivative of Ψ at a point. This comes at the price of altering the higher-order terms in the Taylor series expansion.

Appendix E

Cancellation of the kernel contributions as $k \to 0$ for ΛCDM without radiation

In this Appendix we show the exact cancellation of the three contributions to the kernel of the effective velocity given in (5.17)-(5.19) for the largest scales in a universe with only matter and Λ. In this case, $y = a/a_{\rm eq} \to \infty$, so we can approximate $D_\Psi(a)$ using

$$D_\Psi(a) \equiv \frac{\Psi_{\rm SH}(a)}{\Psi_{\rm SH,i}} = \frac{9}{10}\left[\frac{5}{2}\Omega_m \frac{E(a)}{a} G(a)\right], \tag{E.1}$$

where $G(a) \equiv \int_0^a da' \left[E(a')a'\right]^{-3}$ and $E(a) = \sqrt{\Omega_m a^{-3} + \Omega_\Lambda}$. Further, the distance along the electron's past light cone to redshift $z = 1/a - 1$, normalized by H_0 is given by

$$\Delta\chi(a) = -\int_{a_e}^{a} da' \frac{1}{E(a')a'^2}. \tag{E.2}$$

Let's begin with the simple Sachs-Wolfe term. Expanding to linear order in k, we obtain

$$\begin{aligned}
\mathcal{K}_{\rm SW} =& 3\left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2}\right) j_1(k\Delta\chi_{\rm dec}) \\
=& 3\left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2}\right) \frac{k\Delta\chi_{\rm dec}}{3} + \mathcal{O}(k^3) \\
=& \left(2D_\Psi(\chi_{\rm dec}) - \frac{3}{2}\right) k\Delta\chi_{\rm dec} + \mathcal{O}(k^3).
\end{aligned} \tag{E.3}$$

Expanding the Bessel functions, the Doppler piece will only have two terms at linear order:

$$\mathcal{K}_{\rm D} = -kD_v(a_e) + kD_v(a_{\rm dec}) + \mathcal{O}(k^3) \tag{E.4}$$

We can simplify D_v as follows

$$\begin{aligned} D_v(a) =& \frac{2a^2H(a)}{\Omega_m}\frac{y}{4+3y}\left[D_\Psi(a)+\frac{dD_\Psi(a)}{d\ln a}\right] \\ =& \frac{2a^2E(a)}{\Omega_m}\frac{1}{3}\left[D_\Psi(a)+a\frac{dD_\Psi(a)}{da}\right] \\ =& \frac{2a^2E(a)}{3\Omega_m}\left[-\frac{3\Omega_m}{2E^2(a)a^3}\left(D_\Psi(a)-\frac{3}{2}\right)\right] \\ =& -\frac{1}{aE(a)}\left(D_\Psi(a)-\frac{3}{2}\right). \end{aligned} \tag{E.5}$$

This allows us to write the Doppler piece of the kernel as

$$\begin{aligned} \mathcal{K}_{\rm D} =& \frac{k}{a_eE(a_e)}\left(D_\Psi(a_e)-\frac{3}{2}\right)-\frac{k}{a_{\rm dec}E(a_{\rm dec})}\left(D_\Psi(a_{\rm dec})-\frac{3}{2}\right)+\mathcal{O}(k^3) \\ =& \frac{3k}{2}\left(\frac{1}{a_{\rm dec}E(a_{\rm dec})}-\frac{1}{a_eE(a_e)}\right)+\frac{9k}{4}\Omega_m\left(\frac{G(a_e)}{a_e^2}-\frac{G(a_{\rm dec})}{a_{\rm dec}^2}\right)+\mathcal{O}(k^3). \end{aligned} \tag{E.6}$$

The ISW term can be shown to exactly cancel the above two contributions. We start by expanding $j_1(k\Delta\chi(a)) \sim (k\Delta\chi(a))/3$ and integrating by parts:

$$\begin{aligned} \mathcal{K}_{\rm ISW} =& 6\int_{a_{\rm dec}}^{a_e} da\frac{dD_\Psi(a)}{da}\, j_1(k\Delta\chi(a)) \\ =& \left[6D_\Psi(a)\frac{k\Delta\chi(a)}{3}\right]_{a=a_{\rm dec}}^{a=a_e}-6\int_{a_{\rm dec}}^{a_e} da\frac{kD_\Psi(a)}{3}\frac{d\Delta\chi(a)}{da}+\mathcal{O}(k^3) \\ =& -2D_\Psi(a_{\rm dec})k\Delta\chi_{\rm dec}+2k\int_{a_{\rm dec}}^{a_e} da\frac{D_\Psi(a)}{a^2E(a)}+\mathcal{O}(k^3) \\ =& -2D_\Psi(a_{\rm dec})k\Delta\chi_{\rm dec}+\frac{9k}{2}\Omega_m\int_{a_{\rm dec}}^{a_e} da\frac{G(a)}{a^3}+\mathcal{O}(k^3). \end{aligned} \tag{E.7}$$

The integral in the second term becomes

$$\begin{aligned} \int_{a_{\rm dec}}^{a_e} da\frac{G(a)}{a^3} =& \left[\frac{-G(a)}{2a^2}\right]_{a=a_{\rm dec}}^{a=a_e}+\int_{a_{\rm dec}}^{a_e}\frac{da}{2a^5E^3(a)} \\ =& -\frac{1}{2}\left(\frac{G(a_e)}{a_e^2}-\frac{G(a_{\rm dec})}{a_{\rm dec}^2}\right)+\frac{1}{3\Omega_m}\int_{a_{\rm dec}}^{a_e}\frac{da}{a}\frac{d}{da}\left[\frac{1}{E(a)}\right] \\ =& -\frac{1}{2}\left(\frac{G(a_e)}{a_e^2}-\frac{G(a_{\rm dec})}{a_{\rm dec}^2}\right)+\frac{1}{3\Omega_m}\left(\left[\frac{1}{aE(a)}\right]_{a=a_{\rm dec}}^{a=a_e}+\int_{a_{\rm dec}}^{a_e}\frac{da}{a^2E(a)}\right) \\ =& -\frac{1}{2}\left(\frac{G(a_e)}{a_e^2}-\frac{G(a_{\rm dec})}{a_{\rm dec}^2}\right)-\frac{1}{3\Omega_m}\left(\frac{1}{a_{\rm dec}E(a_{\rm dec})}-\frac{1}{a_eE(a_e)}+\Delta\chi_{\rm dec}\right). \end{aligned} \tag{E.8}$$

Inserting this into (E.7) gives the final result for the ISW contribution

$$\begin{aligned} \mathcal{K}_{\rm ISW} =& -2D_\Psi(a_{\rm dec})k\Delta\chi_{\rm dec}-\frac{9k}{4}\Omega_m\left(\frac{G(a_e)}{a_e^2}-\frac{G(a_{\rm dec})}{a_{\rm dec}^2}\right) \\ & -\frac{3k}{2}\left(\frac{1}{a_{\rm dec}E(a_{\rm dec})}-\frac{1}{a_eE(a_e)}+\Delta\chi_{\rm dec}\right)+\mathcal{O}(k^3) \end{aligned} \tag{E.9}$$

It is now clear that adding equations (E.3), (E.6) and (E.9) gives

$$\mathcal{K}_{\rm SW}+\mathcal{K}_{\rm D}+\mathcal{K}_{\rm ISW}=0+\mathcal{O}(k^3). \tag{E.10}$$

When radiation is included, this cancellation still holds as shown numerically in figure 5.2.

Appendix F

Random Gaussian fields

Following Ref. [238], we generate realizations of the primordial potential $\Psi(\mathbf{x})$ in a four-step process. Given a spatial grid of size L^3 with N^3 positions $\mathbf{x}(\mathbf{m}) = L\mathbf{m}/N$, labeled by the integer triplet $\mathbf{m}$ with components $m_i \in [0, N)$:

1. Define a field $\xi(\mathbf{m})$ that lives on the grid. Draw $\xi(\mathbf{m})$ at each $\mathbf{m}$ from an independent Gaussian probability distribution function with variance N^3.

2. Fourier transform to get

$$\xi(\kappa) = N^{-3} \sum_{\mathbf{m}} \exp\left[-\frac{2\pi i}{N}\kappa \cdot \mathbf{m}\right], \tag{F.1}$$

   where $\kappa \equiv kL/(2\pi)$ is the dimensionless wavenumber.

3. Multiply $\xi(\kappa)$ by

$$F(k) \equiv \left[\left(\frac{2\pi}{L}\right)^3 P_\Psi(k)\right]^{1/2}, \tag{F.2}$$

   where $P_\Psi(k)$ is the Gaussian primordial power spectrum of ΛCDM. Here, we use parameters for the amplitude and spectral index consistent with Planck [8] ($A_s = 2.2 \times 10^{-9}$, $n_s = 0.96$).

4. Inverse Fourier transform to obtain a random field with the correct correlation properties in real and Fourier space:

$$\Psi_i(\mathbf{m}) = \sum_{\kappa} F(k)\xi(\kappa) \exp\left[\frac{2\pi i}{N}\kappa \cdot \mathbf{m}\right]. \tag{F.3}$$